Draft BHA Guidance for an Abbreviated Statement of Work (SoW)
Baseline and Endline Data Collection for Emergency Programs (Revised May 2020)

Applications for emergency activities that are **3 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 implementing 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 onset of the activity.

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

If the applicant plans to conduct an evaluation the applicant may either develop two separate SoW - one for baseline and endline data collection and the second for the evaluation - or combine them into one SoW. Please refer to BHA's Guidance for Evaluation SoW for Emergency Programs.

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. A full baseline protocol must be submitted as an annex to the baseline report*. It is recommended that the successful applicant develop and submit a survey protocol after receiving the award for BHA review.*

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 Emergency P&G 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 to program participants, the general population of the communities being served, or via census:

a) *Beneficiary-based survey*: Beneficiary-based surveys (BBSs) are conducted among the target population that will receive goods and services from an activity. The BBS must use a probability sample. The awardee may conduct a probability-based sample of beneficiaries either during registration or at a later time before implementation of interventions. If the activity uses rolling beneficiary registration or cohorts, baseline data collection should be collected on a rolling basis or for each cohort.

b) *Population-based survey*: In cases where an awardee is implementing community-level interventions such as repairing or installing water points that will benefit the community, a population-based survey (PBS) may be conducted. A PBS is a probability survey that is designed to be representative of all households in the area of implementation. Because population-based surveys can be resource-intensive, partners should only use them when specific indicators need to be measured at the population level as outlined in the PIRS.

c) *Census*: Census is a form of data collection in which all members in the target group or community are included as respondents in the survey. A census may be appropriate if the applicant collects baseline data during the beneficiary registration process and opts to collect data from *all* beneficiaries. A census is likely to require considerable time and financial resources, therefore, applicants should only consider a census when it is more efficient compared to a sample survey.

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 the PIRS and Emergency P&G 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 the Policy and Guidance document 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 Emergency P&G and above), 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.

Please note that the baseline and endline data collection is typically collected, stored, and used such that the confidentiality of participants can be ensured. This means that personally identifiable information (PII) is collected and then kept private. The applicant/ awardee should develop and strictly follow the protocol to decide the level of dataset access. When submitting the dataset to DDL, the partner can designate if the dataset should be made public without restrictions, see ADS 579 for further details.

In order to submit machine-readable quantitative or qualitative data to the 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. BHA offers an example of an Informed Consent in Emergency P&G.

ANNEX A: DETAILED GUIDANCE FOR PROBABILITY-BASED BASELINE/ENDLINE SURVEYS

BHA requires the collection and analysis of quantitative survey data at baseline and endline as appropriate. Quantitative baseline/endline surveys must utilize the same data collection instruments, level of statistical precision, and statistical power. The survey design should be designed to detect statistically significant changes in estimates from baseline to endline for key indicators, described below.

For food security indicators, the quantitative baseline/endline surveys should be conducted in the same season when possible to ensure comparability of data unless the endline data will be collected while transfers are still ongoing. In that case, seasonality will likely have only a marginal effect on food security indicators.

DETERMINING NUMBER OF SAMPLING FRAMES

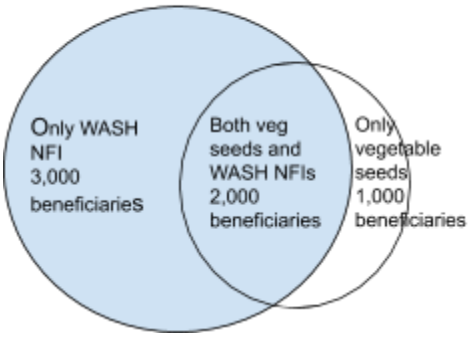
In developing sampling frames for the baseline survey, an Activity needs to take into account its targeting strategy. If an Activity targets households and aims to see a change primarily in household level indicators, even if the activity targets a subset of beneficiaries with specific interventions, the baseline survey will need to develop one sampling frame. While the sample

Figure: 1

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 when one sampling frame will be adequate. Assume an activity targets 5,000 households for food assistance, and targets 3,000 households (a subset) with nutrition-specific interventions.

Assume, another activity (as illustrated in Figure 2) targets 5,000 beneficiaries to distribute WASH NFIs, and also distribute high quality vegetable seed to 3,000 beneficiaries. Note that 2,000 beneficiaries who will receive vegetable seeds will also receive WASH NFIs. However, 1,000 beneficiaries will only receive vegetable seeds. In this example, the baseline survey needs to develop three sampling frames - one with the 3,000 beneficiaries who receive only WASH NFIs (sampling frame 1), one with the 1,000 beneficiaries who will receive only vegetable seeds (sampling frame 2), and the third sampling frame with the 2,000 beneficiaries who will receive both vegetable seeds and WASH NFIs (sampling frame 3).

Figure 2



Using the example above, if the minimum estimated sample size for the beneficiaries who receive WASH NFIs is 660, and minimum sample size for beneficiaries who receive vegetable seeds is also 660, the survey can proportionally distribute the sample as described below.

Since 3,000 beneficiaries are 60% of total beneficiaries who will receive WASH NFIs, the baseline survey should randomly select 396 (60% of 660) sample from sampling frame 1. The remaining 264 sample should be randomly drawn from the sampling frame 3.

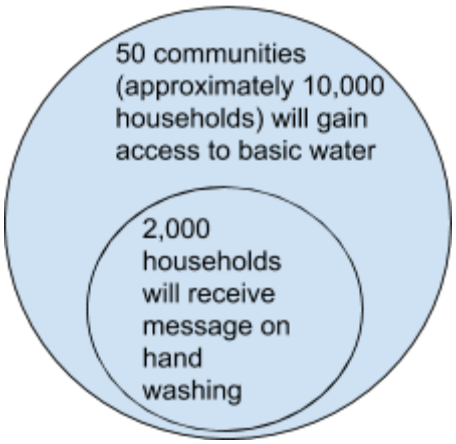
For the vegetable seeds indicator, the estimated sample size is also 660. 1,000 beneficiaries (33% of 3,000) will receive vegetable seeds only. The baseline survey should randomly select 220 beneficiaries (33% of 660) from the sampling frame 2. The remaining 440 sample should be drawn from sampling frame 3.

To maximize efficiency, since 440 households will be randomly selected from sampling frame 3 for the seeds indicator, these 440 households can be used to cover the 264 households needed from sampling frame 3 for the WASH indicators. Therefore, instead of sampling a total of 1,320 households (660+660), a sample size of 1,056 households (440+396+220) will be sufficient.

	Number of Units on Sampling Frame	Sample Size Requirement (WASH)	Sample Size Requirement (SEEDS)	Joint Sample Size
Overlap	**2,000**	**264**	**440**	**440**
WASH ONLY	**3,000**	**396**		**396**
SEEDS ONLY	**1,000**		**220**	**220**
TOTAL	**6,000**	**660**	**660**	**1,056**

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:

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

Since anyone who lives in the targeted 50 communities, can access water from the borehole pumps to be repaired, the activity needs to design a population-based survey to estimate the baseline for the indicator. From a logistical consideration, it is better to design a two-stage cluster survey for basic water services. The minimum estimated sample size for this indicator for a two-stage cluster survey is 660 households.

For the hand washing at critical times indicator, the activity only targets 2,000 households who live in 20 villages. Since these households are selected based on a beneficiary selection criteria, and the activity has a list of the households, the activity should design a beneficiary based household survey with single stage simple random sample design. The estimated minimum sample size for a single stage simple random sample would be 340 households. It is simpler and cleaner for the activity to design two separate surveys with two sampling frames and separate questionnaires.

Interventions /indicators	Potential survey design considerations	Whether population or beneficiary based	Sampling units	Sample Size Requirement
Access to water	Two stage	Population based	50 clusters in stage 1; All households in 30 clusters in stage 2	660
Hand washing	Single stage	Beneficiary based	2,000 beneficiary households	340

TARGET GROUP BY INTERVENTION

Based on the targeting strategy, a baseline survey design that would require multiple sampling frames must organize the target group that will receive a similar set of interventions. Applicants may want to use the table below to assist with identifying sample 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 sampling frame. Since conducting surveys with multiple sampling frames is complex, it is highly recommended to either hire a consultant and/or seek assistance from BHA M&E Advisor.

Intervention [e.g. Cash transfer, IYCF]	Target Group [e.g. all households, women of reproductive age]	Target Number	Assumption/Calculation

Sampling Strategy

Each emergency activity is required to 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. The beneficiary register is the best source of information to construct a sampling frame because it should perfectly reflect the target population.

The sampling frame should include the following key elements:

- Unique household identification number or unique beneficiary identification number (depending on the 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
- Target criteria met

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. In developing the sampling strategy for the baseline and endline, the applicant should choose from the following two strategies.

One-stage Simple Random Sample (SRS) *(recommended when possible)*

If a list of all participants or households is available, and the logistical burden of data collection is reasonable, BHA recommends a **one-stage simple random sampling** 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, producing sampling weights is 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 (i.e. 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, and then draw the sample using multiple stages.

Two-stage Cluster Sampling

Cluster sample designs are typically used in surveys when the logistical costs of data collection using a one-stage SRS are high because the communities 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, primary sampling units (i.e. 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. Therefore, cluster sampling requires the use of a design effect greater than one. When the true design effect is not known, please use 2 as the design effect. This will result in a sample size that is twice as large as estimated using SRS. To minimize intra-cluster correlation, BHA recommends that applicants should 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 primary sampling units (psu).

1) 22 clusters x 30 psu = 660
2) 33 clusters x 20 psu = 660
3) 44 cluster x 15 psu = 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. It may also increase the logistical burden hence cost. Therefore, the activity must carefully consider the cost and advantage to determine the sampling options.

- **Probability Proportional to Size (PPS)**

For a two-stage cluster sample design, it is recommended to use PPS for the first stage of sampling. The PPS method ensures that clusters with more households have a higher probability of selection compared to clusters with fewer households. As a result, data generated using a PPS method do not need to be weighted, simplifying the analysis. To use PPS, a measure of size is needed for each cluster on the sampling frame, i.e. number of households in the cluster, population of the cluster, etc.

Sample Size Calculation

For Indicators Expressed as Proportion

BHA recommends using **a minimum number of 339 respondents** for one-stage (SRS), and 678 respondents for a two- stage sampling design for indicators expressed as proportion. Ideally an applicant must calculate sample sizes separately for each key indicator and take the largest sample size. The formula for calculating sample sizes for indicators expressed as proportion is:

$$n_{initial} = D_{est}\left\lceil \frac{Z_{\text{"\% of households with acceptable FCS score"}}\sqrt{2\underline{P}(1-\underline{P})}+Z_{1-\beta}\sqrt{P_{1,\,est}\,(1-P_{1.est})+P_{2,\,est}\left(1-P_{2,\,est}\right)}}{\delta}\right\rceil^{2}$$

Where

$n_{initial}$ = is the initial sample size required by the surveys for each of the two time points

$\delta = P_{1,\,est} - P_{2,\,est}$ = minimum effect size to be achieved over the time frame specified by the two surveys

$P_{1,\,est}$ = represents a survey estimate of the true population proportion P_1 at baseline [If such an estimate is not available from prior surveys, please use 0.5]

$P_{2,\,est}$ = represents a survey estimate of the true population proportion P_2 at endline

$$\underline{P} = \frac{P_{1,\,est} + P_{2,\,est}}{2}$$

$Z_{1-\infty}$ is the value from the normal probability distribution corresponding to a confidence level $1-\alpha$. For $1-\alpha$=0.95, the corresponding value is $Z_{0.95}$ = 1.64.

$Z_{1-\beta}$ is the value from the normal probability distribution corresponding to a power level of $1-\beta$. For $1-\beta$ = 0.80, the corresponding value is $Z_{0.80}$ =0.84.

D_{est} is the estimated design effect (DEFF) of the survey.

In this example we used "% of households with acceptable FCS score to estimate sample size but this sample size is relevant for all indicators expressed as proportion when we expect and ten percentage point change between baseline and endline. If the expected change between baseline and endline is smaller, we will need a larger sample size.:		
	Single stage SRS	Two stage
$P_{1,\,est}$	50% (0.5)	50% (0.5)
$P_{2,\,est}$	40% (0.4)	40% (0.4)
$Z_{1-\infty}$	95% (1.64)	95% (1.64)
$Z_{1-\beta}$	80% (0.84)	80% (0.84)
D_{est}	1	2
$n_{initial}$	305	610
Non-response adjustments	10%	10%
n_{final}	339	678

Since most indicators presented in table A are exp[ressed as proportion or percentage, the abbreviated guidance does not present guidance for indicators expressed as mean and total. If an applicant identifies key outcome indicators expressed as mean or total for a survey, please contact the BHA M&E Advisor responsible to provide backstopping support to the award.

Sampling Frame

Describe the lists from which primary sampling units (i.e. participants or households) will ultimately be selected.

If a two-stage sampling design is chosen please describe the lists of geographic units that will be used (e.g., villages/communities or camps). It should be stated that a complete list of implementation clusters (villages or communities) will be provided to the survey team.

Selection of Primary Sampling Units, and Clustering at Each Stage of Sampling

This section describes how the participants will be selected if the applicant proposes a two-stage cluster survey, this section should describe how the clustering will be done and how the clusters and primary sampling (direct participants) units will be selected.

ANALYSIS PLAN

For quantitative surveys, describe how the baseline and endline data will be statistically compared, as appropriate. 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, the applicant/awardee should test the difference between baseline and endline FCS raw score as well as "% of households with acceptable FCS score". For HHS, the applicant/awardee 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".

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 raw 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 raw score	Two-sample t-test; One-sample t-test be used when the baseline data was collected through census