Feed the Future

Survey Implementation

Document

Sample Design

Zone of Influence Midline and Resilience Focus Zone Baseline Survey

**Kenya 2023**

*Feed the Future Phase Two Zone of Influence Midline and Resilience Focus Zone Baseline Indicator Assessment*

Revised: March 17, 2023

Submitted: March 7, 2023

This publication was prepared for review by the United States Agency for International Development. It was prepared by ICF for the United States Agency for International Development, USAID Contract Number GS00Q14OADU209/Order Number 7200AA21M00009.

**TABLE OF CONTENTS**

[Abbreviations 3](#_heading=h.23ckvvd)

[1.](#_heading=h.ihv636) Background 4

[2.](#_heading=h.32hioqz) Purpose of this document 5

[3.](#_heading=h.1hmsyys) Target population 5

[4.](#_heading=h.41mghml) Sampling frame 7

[4.1](#_heading=h.3fwokq0) Preparation of the first-stage sampling frame 8

[5.](#_heading=h.1v1yuxt) Sample size determination 9

[5.1](#_heading=h.4f1mdlm) Calculation of initial main survey sample size 10

[5.2](#_heading=h.2u6wntf) Calculation of the final main survey sample size 11

[5.2.1](#_heading=h.19c6y18) Final main survey sample size for the P2-ZOI and RFZ 13

[6.](#_heading=h.28h4qwu) Stratification, sample allocation, and sample selection 14

[6.1](#_heading=h.nmf14n) Stratification 14

[6.1.1](#_heading=h.37m2jsg) Explicit stratification 14

[6.1.2](#_heading=h.1mrcu09) Implicit stratification 14

[6.2](#_heading=h.111kx3o) Sample allocation 15

[6.3](#_heading=h.206ipza) Sample selection 17

[6.3.1](#_heading=h.4k668n3) First-stage sampling of EAs 17

[6.3.2](#_heading=h.2zbgiuw) Preparation of the second- and third-stage sampling frames: Listing 19

[6.3.3](#_heading=h.1egqt2p) Potential second-stage sampling of segmented EAs 19

[6.3.4](#_heading=h.2dlolyb) Third-stage sampling of households within sampled EAs or segments 20

[6.3.5](#_heading=h.3cqmetx) Fourth-stage sampling of individuals within sampled households 20

[7.](#_heading=h.2r0uhxc) Reserve sample 22

[8.](#_heading=h.1664s55) Sample weighting 23

[8.1](#_heading=h.3q5sasy) Calculating probabilities of selection and design weights 24

[8.2](#_heading=h.25b2l0r) Adjustments to the design weights 25

[8.3](#_heading=h.kgcv8k) Final sampling weights 25

[9.](#_heading=h.34g0dwd) Calculating measures of precision 29

[10.](#_heading=h.43ky6rz) References 30

[Annex: List of Sub-zones and counties comprising the Kenya phase two zone of influence and resilience focus zone 31](#_heading=h.xvir7l)

**LIST OF FIGURES**

[Figure 1: Map of Kenya: P2-ZOI and RFZ Counties 6](#_heading=h.46r0co2)

**LIST OF TABLES**

[Table 1: Distribution of the RFZ and P2-ZOI Counties for the Midline Survey 7](#_heading=h.2lwamvv)

[Table 2: Calculation of Initial Sample Size for the P2-ZOI Midline Survey Indicators 10](#_heading=h.3ygebqi)

[Table 3: Calculation of Final Sample Size for the P2-ZOI Midline Survey Indicators 12](#_heading=h.sqyw64)

[Table 4: Minimum Sample Size Required for the EBF Indicator 13](#_heading=h.1rvwp1q)

[Table 5: Distribution of the Midline Survey EAs per Stratum 15](#_heading=h.1jlao46)

[Table 6: Proportional Allocation of Final Sample Size across Strata 16](#_heading=h.2iq8gzs)

[Table 7: Distribution of the Main Sample EAs per Stratum 18](#_heading=h.1x0gk37)

[Table 8: Individual-level Sampling Groups 21](#_heading=h.4bvk7pj)

[Table 9: Distribution of the Main and Reserve Sample EAs per Stratum 22](#_heading=h.1opuj5n)

[Table 10: List of Directly Collected or Imputed Indicators and Their Sampling Weights 26](#_heading=h.1pxezwc)

[Table 11: List of Additional Topics and Their Sampling Weights 28](#_heading=h.49x2ik5)

# Abbreviations

| A-WEAI | Abbreviated Women’s Empowerment in Agriculture Index |
| --- | --- |
| DHS | Demographic and Health Surveys |
| EA | enumeration area |
| EBF | exclusive breastfeeding |
| FIES | Food Insecurity Experience Scale |
| HWISE-4 | Brief Water Insecurity Experiences Scale |
| KEA  MOS | Kenya and East Africa  measure of size |
| PBS | population-based survey |
| PPS  PSU | probability proportional to size  primary sampling unit |
| RFS | Bureau for Resilience and Food Security |
| RFZ | Resilience Focus Zone |
| SMRFS | Surveys for Monitoring in Resilience and Food Security |
| USAID | United States Agency for International Development |
| ZOI | Zone of Influence |

# Background

As the U.S. Government’s global hunger and food security initiative, Feed the Future seeks to reduce poverty, hunger, and undernutrition among women and children, and to increase income, resilience, women’s empowerment, dietary diversity, and appropriate feeding practices, and improve hygienic environments. Program efforts are designed to impact the population in Zones of Influence (ZOIs) and, where they overlap with ZOIs, Resilience Focus Zones (RFZs) in Feed the Future target countries. Progress in achieving Feed the Future’s objectives is tracked using population-based performance indicators collected at baseline then periodically thereafter.

Programming in RFZs is designed to increase the ability of people, households, communities, countries, and systems to mitigate, adapt to, and recover from shocks and stresses in a manner that reduces chronic vulnerability and facilitates inclusive growth. One measure of the success of programming in the RFZ is a reduction in the need for humanitarian food assistance.

Kenya is one of the target countries of phase two of the Feed the Future program. A midline indicator assessment will be conducted in the Kenya phase two ZOI (P2-ZOI), which will also serve as a baseline indicator assessment in the RFZ, in May-July 2023. The Feed the Future 2023 Kenya P2-ZOI Midline and RFZ Baseline Indicator Assessment is being implemented under the Surveys for Monitoring in Resilience and Food Security (SMRFS) task order. The objective of SMRFS is to provide the United States Agency for International Development’s (USAID) Bureau for Resilience and Food Security (RFS) and Feed the Future target, resilience focus, and Bureau for Humanitarian Assistance Resilience Food Security Activity countries with high-quality population-based survey (PBS) data using an approach that reflects state-of-the-art methods and is standardized across surveys. Under SMRFS, ICF will support the design and implementation of PBS data collection, analysis, and reporting. The data from the SMRFS project will support the monitoring and understanding of progress under important U.S. Government strategies, such as the Multi-Sectoral Nutrition Strategy, the Global Water Strategy, and the Feed the Future Global Food Security Strategy.

The purpose of the P2-ZOI Midline and RFZ Baseline Indicator Assessment is to provide U.S. Government interagency partners, the USAID Bureau for Resilience and Food Security, USAID/Kenya and East Africa (USAID/KEA), the Government of Kenya and development partners with information about short-term progress of P2-ZOI indicators and establish baseline values for RFZ indicators in Kenya. The Midline Indicator Assessment will collect data through the P2-ZOI Midline and RFZ Baseline Survey, hereafter referred to as the Midline Survey, which comprises a main and parallel survey and an RFZ survey. This is the first assessment of the Feed the Future P2-ZOI indicators since the baseline indicator values were established in 2019. This Midline Survey is not designed to detect statistically significant changes in indicator values since the baseline; however, it will be possible to detect statistically significant differences if the difference between baseline and midline values is large enough.

# Purpose of this document

This document provides details on the steps undertaken in development of the sampling frame and sample design for the Midline Survey. These activities were defined by the sample requirements for the survey, following the guidelines as outlined in the *Guidance on the Feed the Future Phase Two Zone of Influence Midline Indicator Assessment* and the *Feed the Future Population-Based Survey Sampling Guide*, including guidance for the RFZ survey as outlined in the scope of work document.[[1]](#footnote-0),[[2]](#footnote-1),[[3]](#footnote-2)

**Parallel survey**   
As part of the P2-ZOI Midline Indicator Assessment, a parallel survey in addition to the main survey (i.e., P2-ZOI Midline Survey) will also be completed as part of this task order to collect consumption expenditure information on a sample of 352 households in the same enumeration areas as the main survey for the P2-ZOI. The parallel survey data will be used to validate a baseline prediction model used for survey-to-survey imputation and provide a back-up method to compute midline estimates for poverty indicators if needed. Specific information about the parallel survey sample is included in the blue text boxes throughout this document.

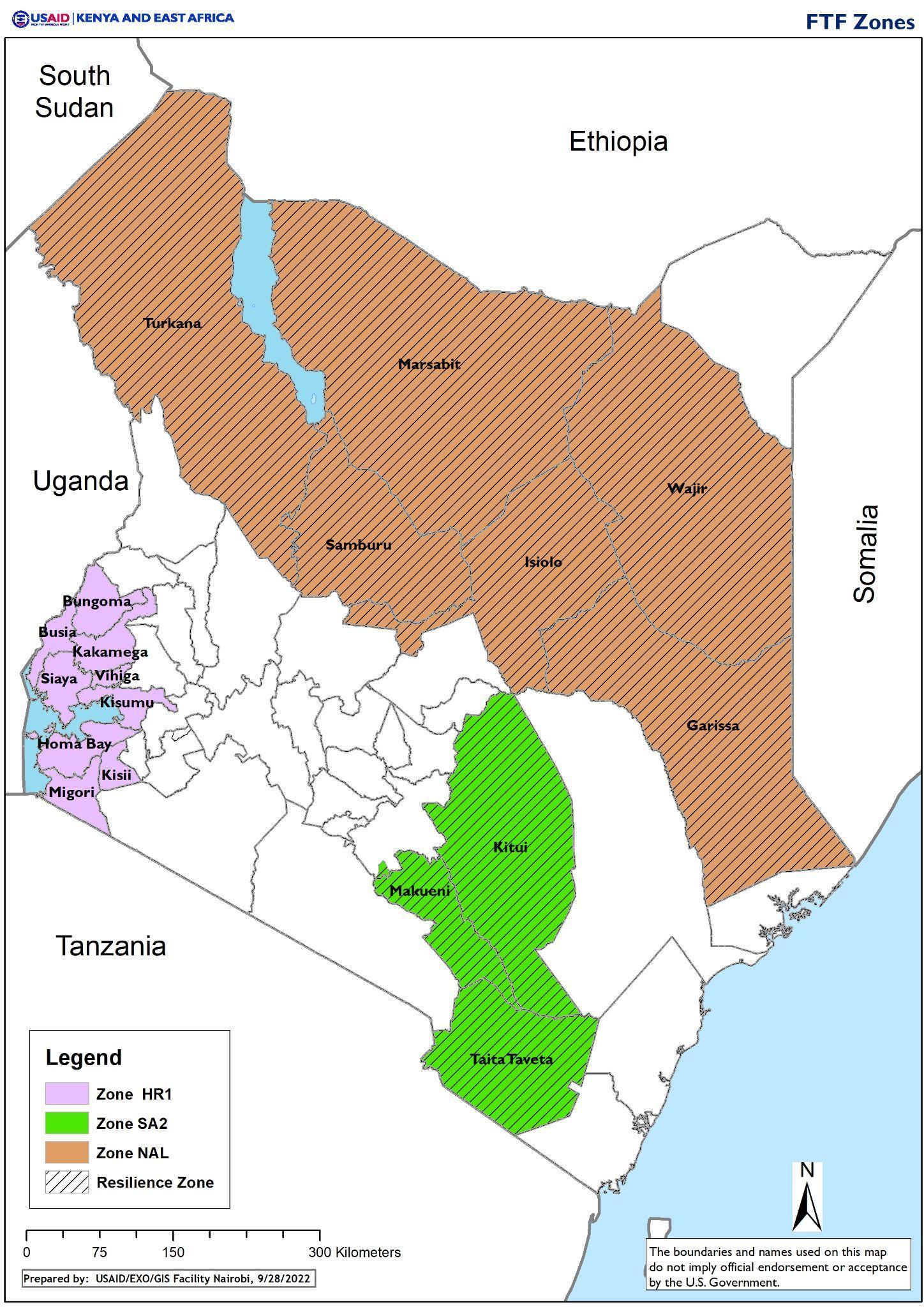
# Target population

The target population for the Midline Survey is all households and persons residing within the counties that are part of the P2-ZOI and RFZ in Kenya. The P2-ZOI covers 18 counties across three sub-zones: a high rainfall area 1 (HR1), a semi-arid area 2 (SA2) and the northern arid lands (NAL). Within the P2-ZOI, USAID/KEA has designated a RFZ. The RFZ is comprised of 9 counties, including all 3 counties in the SA2 sub-zone and all 6 counties in the NAL sub-zone (see Figure 1).[[4]](#footnote-3)

**Table 1** shows the distribution of enumeration areas (EAs) and households based on the 2019 Kenya Population and Housing Census data across these 18 counties by sub-zone.

The HR1 sub-zone has the largest share of the P2-ZOI households, with 9 counties, making up approximately 68 percent of the P2-ZOI households. The SA2 sub-zone with 3 counties and the NAL sub-zone with 6 counties, each make up approximately 16 percent of the P2-ZOI households.

**Figure 1: Map of Kenya: P2-ZOI and RFZ Counties**



The RFZ comprises the 3 counties in the SA2 sub-zone, making up approximately 51 percent of the RFZ households, and the 6 counties in the NAL sub-zone, making up 49 percent of the RFZ households.

**Parallel survey target population**  
The target population for the parallel survey is the same as the target population for the P2-ZOI Midline Indicator Assessment (i.e., all households residing within the 18 counties that are part of the P2-ZOI in Kenya).

**Table 1: Distribution of the RFZ and P2-ZOI Counties for the Midline Survey**

| **Sub-zone** | **County** | **EAs** | **Households** |
| --- | --- | --- | --- |
| High Rainfall Area 1 (HR1) | Bungoma | 3,732 | 357,029 |
| Busia | 2,063 | 197,830 |
| Homa Bay | 2,888 | 260,385 |
| Kakamega | 4,605 | 433,000 |
| Kisii | 3,526 | 307,254 |
| Kisumu | 2,933 | 296,349 |
| Migori | 2,585 | 238,334 |
| Siaya | 2,676 | 249,428 |
| Vihiga | 1,401 | 143,284 |
| **Total (HR1)** | | **26,409** | **2,482,893** |
| **Northern Arid Lands (NAL)** | Garissa\* | 1,312 | 109,085 |
| Isiolo\* | 505 | 52,775 |
| Marsabit\* | 866 | 76,833 |
| Samburu\* | 697 | 64,219 |
| Turkana\* | 1,694 | 146,091 |
| Wajir\* | 1,297 | 126,318 |
| **Total (NAL)** | | **6,371** | **575,321** |
| Semi-Arid Area 2 (SA2) | Kitui\* | 3,731 | 261,664 |
| Makueni\* | 2,778 | 243,931 |
| Taita/Taveta\* | 1,009 | 93,981 |
| **Total (SA2)** | | **7,518** | **599,576** |
| **Total (RFZ)** | | **13,889** | **1,174,897** |
| **Total (P2-ZOI)** | | **40,298** | **3,657,790** |

\* Denotes counties that are part of the RFZ

# Sampling frame

The Feed the Future PBS sampling guide recommends the use of a stratified, multi-stage cluster sample design to obtain a representative random sample for surveys.[[5]](#footnote-4) The multiple stages of sample selection include a sample of primary sampling units (PSUs) at the first stage, as well as segments (if applicable), households, and individuals at subsequent stages. To select a probability sample that is representative of the population, a valid frame of the sampling units at each stage of sample selection is required.

This section discusses the sampling frames that are used at each stage of sampling for the Midline Survey in Kenya. A high-quality sampling frame is critical to the successful implementation of any PBS.[[6]](#footnote-5) It comprises a complete list of all sampling units that covers the target population (i.e., clusters/EAs/wards, segments, households, or individuals) from which a representative sample can randomly be selected at each stage of the survey. Without such frames, it is difficult to ensure a representative sample for the survey.

The Feed the Future PBS sampling guide recommends the use of census EAs as the first-stage sampling units/clusters for PBSs. The recommended sampling frame at the first stage is therefore a census EA frame that covers the target population. For the Midline Survey in Kenya, the sampling frame was based on the 2019 Kenya Population and Housing Census EAs. The EAs that fall within the 18 counties making up the RFZ and P2-ZOI constitute the sampling frame for the Midline Survey.

At the second stage, there is a possibility of sampling segments within some EAs, depending on the size of the EAs found during listing operations. EAs that are found to be too large to conduct listing and fieldwork operations feasibly will be subject to segmentation, following the guidelines in the Feed the Future PBS sampling guide.[[7]](#footnote-6) Segmentation is a process whereby EAs are broken down further or “segmented” into smaller geographic areas (based on a set criterion), and an additional stage of sampling of these segments is implemented. It should be noted that segmentation would only be implemented in EAs that are found to meet the requirements for segmentation, and the identification of these EAs will only be done during the initial stages of the listing process. Therefore, for the Kenya Midline Survey, a sampled cluster is an EA or a segment of an EA.

Within the selected clusters, a complete and comprehensive list of all households will be created through a household listing process that occurs before the fieldwork for the survey. This list of households in each cluster will constitute the sampling frame for selecting a representative random sample of households for the survey.

At the final stage of sampling, the frame consists of individuals found within the sampled households from the previous stage. This frame is established through a household roster, which is a listing of all household members, along with associated information such as demographics and relationship to other members within the household.[[8]](#footnote-7) Interviewers generate the household roster during the main data collection, after sampled households have been located and contact has been established with a responsible adult within these households.

Currently, the first-stage sampling of EAs has been completed for the Kenya Midline Survey (see Section 6.3.1 for more details on how the first-stage sample was selected). The household listing operations will now begin within these sampled EAs (Sections 6.3.2 and 6.3.3 provide more details on the procedures for segmentation [if applicable] and household listing). When the household listing operations are completed, the household sampling frame for this stage of the sample design will be available for sampling of households (see Section 6.3.4 for more details on the proposed procedures for household sampling). Finally, this is followed by the frame for sampling of eligible individuals within the sampled households (see Section 6.3.5 for more details on the proposed procedures for sampling of individuals within households).

At this point in the implementation of the survey, only the first-stage sampling frame and EA sample have been completed. The remainder of this section discusses in more detail the steps undertaken in preparing the first-stage sampling frame.

## Preparation of the first-stage sampling frame

The key characteristic of a high-quality sampling frame is that it provides a comprehensive, complete, and up-to-date set of sampling units that cover the entire target population. In general, for the first-stage frame to be considered “complete,” it should consist of an exclusive (i.e., with no duplicates) and exhaustive (i.e., with no missing units) set of all EAs covering the 18 counties making up the RFZ and P2-ZOI.

For the first-stage sampling frame to be considered comprehensive, it should include, at a minimum, a unique identifier for each EA (such as a unique EA code), as well as information on the administrative hierarchy and geography in the country. In addition, any auxiliary information that can be used during the sample design will be useful (e.g., number of households in the EA, number of persons in the EA). As mentioned previously, the EAs of the 2019 Kenya Population and Housing Census were used to derive a first-stage sampling frame for the Midline Survey. The frame file provided included an EA code, information on the number of households per EA (based on the 2019 census counts), and the administrative hierarchy and geographic residence type defining each EA.

The use of the census EAs in defining the first-stage frame for the survey is ideal, because the EAs ensure that there is complete coverage of the 18 counties constituting the RFZ and P2-ZOI in Kenya (with no overlaps). In addition, the information collected on the population during the census can easily be aggregated at the EA level to assist in defining an appropriate sample design. The information on the frame file was checked to ensure that all necessary variables were complete and had valid values to enable the sample design for the survey. Overall, 40,298 EAs fall within the 18 counties that are part of the P2-ZOI in Kenya, with 13,889 of these EAs falling within the 9 counties that make up the RFZ.

**Parallel survey first-stage sampling frame**  
The target population for the parallel survey is the same as for the P2-ZOI main survey, and therefore the first-stage sampling frame is also the same (i.e., the EAs that fall within the 18 counties that constitute the P2-ZOI in Kenya).

# Sample size determination

The main aim of the Midline Indicator Assessment is to provide sample-weighted estimates of the P2-ZOI and RFZ population-based indicators (including their standard errors and confidence intervals) to enable performance monitoring between the Kenya Baseline and Midline Surveys. This requires a descriptive PBS with a sample size that is adequate to enable the calculation of reliable indicator estimates at a single point in time (i.e., midline). Although the Midline Survey is not designed to detect statistically significant differences in indicator values since the Baseline Survey, it will still be possible to detect statistically significant differences if the difference between baseline and midline values is large enough.

The sample size was determined following the guidance developed by RFS for P2-ZOI Midline Indicator Assessments and the RFZ.[[9]](#footnote-8) As per the guidance, the required sample sizes for the following directly collected and computed ZOI midline indicators was determined:[[10]](#footnote-9)

* Prevalence of moderate and severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)
* Percent of households below the comparative threshold for the poorest quintile of the Asset-Based Comparative Wealth Index
* Percent of households participating in group-based savings, micro-finance, or lending programs
* Percent of households with access to a basic sanitation service
* Percent of households with soap and water at a handwashing station on premises
* Percent of households that believe local government will respond effectively to future shocks and stresses
* Percent of women of reproductive age consuming a diet of minimum diversity
* Prevalence of exclusive breastfeeding (EBF) of children under six months of age
* Percent of producers who have applied targeted improved management practices or

technologies

* Ability to recover from shocks and stresses index
* Index of social capital at the household level

Given the global coronavirus pandemic shock and the impact it has had on poverty and nutrition in low-income countries, determining the midline sample size based on pre-pandemic assumptions of improvement in indicator values from baseline may not be realistic. Preventing backsliding from baseline levels is considered a positive result under these circumstances. Therefore, the monitoring and evaluation guidance on the Feed the Future P2-ZOI Midline Indicator Assessment allows missions to use baseline indicator values when computing the required midline sample size, except for the indicator *Prevalence of exclusive breastfeeding of children under six months of age,* for which the sample size was determined based on having a minimum of 70 children under 6 months of age in the sample to ensure statistically valid estimates of exclusive breastfeeding rates by sex for the population.

## Calculation of initial main survey sample size

The Midline Survey is a descriptive PBS, with the purpose to provide midline point estimates for the P2-ZOI and RFZ indicators at an acceptable level of statistical accuracy. The initial required sample sizes were computed for the directly collected and computed indicators listed in the previous section. The initial sample size required for the main survey was determined based on an acceptable margin of error for the estimated level and precision of each of the directly collected and computed P2-ZOI indicators.

**Table 2** presents the calculation of the initial sample size for the main survey based on the input parameters for the indicators, as given in the table using the formula that follows.

**Table 2: Calculation of Initial Sample Size for the P2-ZOI Midline Survey Indicators**

| **Indicator type** | **Indicator** |  |  |  | **\*** |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Percentage/proportion | FIES | 72.7 | 5% | 1.96 | n/a | 5.0 | 1,525 |
| Percent households  (Comparative Wealth Index) | 40.5 | 5% | 1.96 | n/a | 8.0 | 2,963 |
| Percent households  (savings programs) | 49.7 | 5% | 1.96 | n/a | 5.9 | 2,267 |
| Percent households  (basic sanitation) | 26.0 | 5% | 1.96 | n/a | 5.6 | 1,656 |
| Percent households  (handwashing) | 8.9 | 5% | 1.96 | n/a | 6.7 | 835 |
| Percent households  (local government) | 60.6 | 5% | 1.96 | n/a | 4.1 | 1,505 |
| Percent women  (dietary diversity) | 25.5 | 5% | 1.96 | n/a | 4.7 | 1,372 |
| Percent producers  (ag practices) | 98.0 | 5% | 1.96 | n/a | 1.4 | 43 |
| Mean | Ability to Recover from Shocks and Stresses Index | 4.0 | 5% | 1.96 | 2.4 | 2.8 | 1,549 |
| Social Capital Index (households) | 50.7 | 5% | 1.96 | 87.8 | 5.4 | 24,885 |

\* Estimated standard deviation for mean indicators.

(1)[[11]](#footnote-10)

Where,

= estimated initial sample size required for the survey based on the indicator.

= the estimated design effect for the indicator in this survey. The estimate was based on the value from the P2-ZOI Baseline Survey for indicators.[[12]](#footnote-11)

= the estimated prevalence or mean of the indicator at the midline. This value is derived from the baseline for this indicator.

= the critical value for the normal probability distribution. The significance level is set at = 0.05, giving a value of = 1.96.

= is the margin of error. The margin of error used for the P2-ZOI Midline Survey is 5 percent (i.e., M = 0.05).

Based on the initial sample size computation (see **Table 2**), the *Index of Social Capital* indicator had the largest initial sample size required of 24,885 households across all indicators.[[13]](#footnote-12)

## Calculation of the final main survey sample size

To derive the final required household sample size, the initial sample size values calculated previously should be adjusted to account for the number of households that need to be contacted to obtain the required number of individuals for the given sub-population for individual level-indicators and anticipated household non-response. Therefore, the final sample size, denoted by , is the product of the initial sample size and the adjustment for the number of households to contact for the given individual sub-population for individual-level indicators (i.e., ) and the adjustment for anticipated household non-response (i.e., ) for each indicator:

(2)

Where,

= estimated final sample size required for the survey based on the indicator

= adjustment for the number of households to contact to obtain the required number of individuals for the given sub-population for an individual-level indicator[[14]](#footnote-13)

= adjustment for anticipated household non-response, which is equal to the inverse of the expected response rate

**Table 3** illustrates the computation of the final required sample size for the Feed the Future P2-ZOI PBS individual and household-level indicators.[[15]](#footnote-14) An expected response rate of 90.2 percent was assumed for the indicators based on the 2019 Kenya P2-ZOI Baseline Survey results.

**Table 3: Calculation of Final Sample Size for the P2-ZOI Midline Survey Indicators**

| **Indicator type** | **Indicator** |  | **Estimated average number of individuals in the sampling group per household** |  | **Expected response rate** |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Percentage/ proportion | FIES\* | 1,525 | n/a | n/a | 90.2% | 1.11 | 1,691 |
| Percent households (Comparative Wealth Index)\* | 2,963 | n/a | n/a | 90.2% | 1.11 | 3,285 |
| Percent households  (savings programs)\* | 2,267 | n/a | n/a | 90.2% | 1.11 | 2,514 |
| Percent households  (basic sanitation) | 1,656 | n/a | n/a | 90.2% | 1.11 | 1,836 |
| Percent households  (handwashing) | 835 | n/a | n/a | 90.2% | 1.11 | 926 |
| Percent households  (local government)\* | 1,505 | n/a | n/a | 90.2% | 1.11 | 1,669 |
| Percent women  (dietary diversity) | 1,372 | 1.19 | 1.20 | 90.2% | 1.11 | 1,828 |
| Percent producers  (ag practices) | 43 | 0.63 | 2.01 | 90.2% | 1.11 | 97 |
| Mean | Ability to Recover from Shocks and Stresses Index\* | 1,549 | n/a | n/a | 90.2% | 1.11 | 1,718 |
| Social Capital Index (households)\* | 24,885 | n/a | n/a | 90.2% | 1.11 | 27,589 |

Overall, the indicator with the largest required final household sample size was the *Index of Social Capital* indicator, requiring 27,589 households. This sample size was compared to the minimum required household sample size for the *Prevalence of exclusive breastfeeding of children under six months of age indicator* (or “EBF indicator”), which was determined using the following formula:

(3)

Where,

= minimum required household sample size for the EBF indicator

= estimated proportion of the P2-ZOI population that is under 6 months of age

= estimated average household size in the P2-ZOI

= inverse of the expected household response rate

**Table 4** shows the parameter values used in calculating the minimum required household sample size for the EBF indicator.

**Table 4: Minimum Sample Size Required for the EBF Indicator**

| **Indicator** |  |  | **Expected response rate** |  |  |
| --- | --- | --- | --- | --- | --- |
| EBF | 5.2 | 0.01 | 90.2 | 1.11 | 1,532 |

As shown in **Table 4,** the EBF indicator requires a sample size of 1,532 households to enable at least an estimated 70 children under 6 months of age to be part of the midline sample.

### 5.2.1 Final main survey sample size for the P2-ZOI and RFZ

Comparing the final required sample sizes across all indicators from **Table 3** and **Table 4,** the *Index of Social Capital* indicator still has the largest required sample size of 27,589 households and therefore meets or exceeds the sample size needs for all indicators. However, given the large, required sample size calculated and its impact on the survey operations, USAID/KEA has determined that the sample size for each of two sub-zones (HR1 and SA2/NAL (i.e., RFZ)) should be capped at 2,200 households in line with Feed the Future monitoring and evaluation guidance on sample size determination for the midline indicator assessment. When applying the required sample size across the two primary strata – HR1 and SA2/NAL - the final sample size for the assessment will therefore be 4,400 households. With a sample take of 25 households per cluster, a total of 176 clusters will be sampled for the main survey, with 88 clusters in the non-RFZ stratum (HR1) and 88 clusters in the RFZ stratum (SA2/NAL).

**Parallel survey sample size**  
The parallel survey sample will be selected as an additional sample of households in the same EAs as the main survey (see Section 6 for description of the parallel survey sample design and selection). To ensure that the minimum required sample size of 300 households for the parallel survey is met, a sample of 2 additional households will be selected for the parallel survey in each of the EAs sampled for the main survey. Therefore, the sample size for the parallel survey will be 352 households (=176 x 2).

# Stratification, sample allocation, and sample selection

The survey is based on a stratified, multi-stage cluster sample design, as recommended by the Feed the Future PBS sampling guide. This section describes the sample design in more detail, specifically with respect to the stratification, sample allocation, and sample selection.

## Stratification

Stratification is the process by which a target population is divided into subgroups (called strata) that have similar characteristics. It is an important step in designing PBSs because it has the effect of reducing the overall variability across the sample, resulting in an increase in the precision of survey estimates. In addition, it allows for separate survey estimates to be produced for strata, such as the case with this survey, which requires separate estimates for the RFZ and P2-ZOI. It also conveniently allows for greater flexibility, in the sense that different sample designs can be applied within each stratum, as well as providing a foundation for allocating the sample. For the Midline Survey, the stratification of the sampling frame will be undertaken in two phases: explicit stratification and implicit stratification.

### 6.1.1 Explicit stratification

The first-level stratification or primary stratification of the sampling frame was undertaken to be able to distinguish between the RFZ areas and non-RFZ areas to be able to produce separate estimates for the RFZ and overall P2-ZOI (see **Table 5**). Within each of these primary strata, further levels of stratification were undertaken to improve sample efficiency and enable the required analyses. In each primary stratum, further stratification by sub-zone (i.e., high rainfall area 1 (HR1), semi-arid 2 (SA2) and northern arid lands (NAL) and within each sub-zone by urban/rural residence was undertaken. Within the NAL sub-zone an additional level of stratification was undertaken to distinguish the Samburu county from the other counties in the NAL sub-zone to enable analysis for baseline comparisons when Samburu county was not part of the RFZ and P2-ZOI. The stratification also ensures that an adequate reserve sample can be allocated within those strata in which it is anticipated that replacement of PSUs will be needed (see Section 7 for more details on the reserve sample). **Table 6** shows the distribution of the EAs per stratum.

The RFZ stratum covers the 9 counties that are part of the RFZ. These counties are spread across the SA2 and NAL sub-zones. The remaining 9 P2-ZOI counties that are not part of the RFZ form the non-RFZ stratum and are in the HR1 sub-zone.

### 6.1.2 Implicit stratification

Implicit stratification is an added measure that can be implemented to improve the efficiency of the sample design prior to sampling without having to create explicit strata. By ordering the frame within each stratum in a logical way and then sampling, gains in precision can be achieved because the ordering has a similar effect as stratification without explicitly creating separate strata.

For the current survey design, within each of the explicit strata (as listed in **Table 5**), the EAs were sorted by county, sub-county, division, location, sub-location village, and EA code before selection of the EAs during the first stage of sampling, to increase the efficiency of the sample design.

**Table 5: Distribution of the Midline Survey EAs per Stratum**

| **Stratum** | | | | **No. of EAs** |
| --- | --- | --- | --- | --- |
| **Primary** | **Sub-zone** | **County** | **Residence** |
| RFZ | Semi-Arid 2 (SA2) | | Urban | 820 |
| Rural | 6,698 |
| Northern Arid Lands (NAL) | Samburu | Urban | 138 |
| Rural | 559 |
| Northern Arid Lands (NAL) | Non-Samburu | Urban | 1,600 |
| Rural | 4,074 |
| **Total (RFZ stratum)** | | | | **13,889** |
| Non-RFZ | High Rainfall 1 (HR1) | | Urban | 4,578 |
| Rural | 21,831 |
| **Total (non-RFZ stratum)** | | | | **26,409** |
| **Overall total** | | | | **40,298** |

## Sample allocation

The main aim of the survey is to produce indicator estimates at the overall level for the P2-ZOI and RFZ separately. As indicated in section 5.2.1, an overall sample of 4,400 households is needed for this assessment with an equal allocation of 2,200 households in the RFZ stratum and non-RFZ stratum (i.e., the HR1 sub-zone). The sample within the RFZ primary stratum will then be equally allocated between the SA2 and NAL sub-strata (i.e., 1,100 households per sub-stratum). Within the HR1 and SA2 sub-strata, the sample will be allocated proportionally by urban/rural residence. In the NAL sub-stratum, the proportional allocation will also account for Samburu and non-Samburu counties in addition to urban/rural residence, to facilitate appropriate analysis and baseline-midline comparisons.

The proportional sample allocation method allocates each of the required samples for the HR1, SA2 and NAL sub-strata proportionally within each stratum using a measure of size (MOS), such as the number of households in each stratum. For this survey, the MOS used was the census household count per stratum. The formula for allocating the final sample size to the various strata using proportional allocation is given by:

(4)

Where:

= portion of the final sample size () to be allocated to stratum

= number of households in stratum (based on the census count)

= number of households across all strata

**Table 6** presents the allocation of the final sample size (=4,400 households) to each stratum. Note, in order to maintain a constant sample size per cluster of 25 households across all sampled clusters, the calculated sample allocations for each stratum using equation (4) was either rounded up or down to a multiple of 25, with at least two clusters being allocated per stratum (i.e., 50 households) to maintain a minimum level of statistical representativeness and be able to estimate measures of precision during data analysis accounting for all strata in the sample. Therefore, the final allocation across the strata may differ slightly from the proportional allocation calculated using equation (4).

**Table 6: Proportional Allocation of Final Sample Size across Strata**

| **Stratum** | | | | **Number of households in stratum ()** | **Allocation of the sample**  **()** |
| --- | --- | --- | --- | --- | --- |
| **Primary** | **Sub-zone** | **County** | **Residence** |
| RFZ | Semi-Arid 2 (SA2) | | Urban | 75,383 | 150 |
| Rural | 524,193 | 950 |
| Northern Arid Lands (NAL) | Samburu | Urban | 12,493 | 50 |
| Rural | 51,726 | 75 |
| Northern Arid Lands (NAL) | Non-Samburu | Urban | 144,074 | 275 |
| Rural | 367,028 | 700 |
| **Total (RFZ stratum)** | | | | **1,174,897** | **2,200** |
| Non-RFZ | High Rainfall 1 (HR1) | | Urban | 423,929 | 375 |
| Rural | 2,058,964 | 1,825 |
| **Total (non-RFZ stratum)** | | | | **2,482,893** | **2,200** |
| **Overall total** | | | | **3,657,790** | **4,400** |

## Sample selection

The sampling process will involve multiple stages of selection to arrive at the final sample to be implemented during fieldwork. This section describes each stage of sampling and how it will be implemented. Currently, the first-stage sample of EAs has been selected (as described in Section 6.3.1), and the listing operation within these sampled EAs will determine the basis for subsequent stages of selection (i.e., procedures described in Sections 6.3.2 and 6.3.3). After the listing operations are completed, the sampling of households will be undertaken (see Section 6.3.4), followed by the sampling of eligible individuals within the sampled households (see Section 6.3.5).

### 6.3.1 First-stage sampling of EAs

After the stratification of the sample frame and sample allocation has been completed, each stratum, denoted by , has its own sample size, , according to the sample allocation scheme used. The next step in the survey design process was to randomly select the sample of EAs from the sampling frame, independently within each stratum.

Systematic probability proportional to size (PPS) sampling was used to randomly select a sample of EAs (called the PSUs) during the first stage of sampling. The advantage of PPS sampling in a multistage design is that it leads to an overall self-weighting scheme (when combined with subsequent stages), resulting in a more efficient sample design for increased precision in estimating population indicators. In general, PPS sampling selects the PSUs according to an MOS that is related to the key indicators of interest. For Feed the Future ZOI PBSs, the recommended MOS to be used is the “total number of households” in each first-stage sampling unit. For the Midline Survey, the 2019 census household counts on the sampling frame were used as the MOS. In addition, Feed the Future ZOI PBSs use a fixed household sample size per PSU to maintain the approximate self-weighting design. The recommended household sample size per PSU for Feed the Future ZOI PBSs (as given in the Feed the Future PBS sampling guide) is within a range of 20 to 30 households per PSU. In most cases, this represents a logistically feasible number of households per PSU in which to conduct interviews without compromising statistical efficiency by inducing a design effect (due to clustering) that is too large. For the Midline Survey, a fixed number of 25 households per PSU will be selected.

Following the recommendations in the Feed the Future PBS sampling guide, a systematic PPS sample of EAs was selected within each stratum based on the allocated stratum sample sizes, (see **Table 7**) and the sample take per EA, with the MOS being the census counts of the number of households within each EA.[[16]](#footnote-15)

During discussions with the local survey implementation partner, it was flagged that some counties within the P2-ZOI and RFZ were anticipated to have accessibility issues due to insecurity and may require replacement EAs. The Feed the Future PBS sampling guide makes provision for the allocation of a reserve sample of EAs that could be used if some of the EAs from the main sample are found to be inaccessible, using a two-phase sampling approach during this first stage of selection. During the first phase of the first stage of sampling, a total number of EAs were sampled from each stratum that constitute both the reserve number of EAs expected to be required as well as the actual number of EAs required for the main sample. At the second phase of the first stage of sampling, the subset of EAs for the reserve sample is then selected from the first phase sample using fractional interval systematic sampling. The reserve EAs will be released during fieldwork as and when required. For more details on the reserve sample EAs, see Section 7. **Table 7** provides the distribution of the EAs for the main sample.

**Table 7: Distribution of the Main Sample EAs per Stratum**

| **Stratum** | | | | **Number of sampled EAs per stratum**  **()** |
| --- | --- | --- | --- | --- |
| **Primary** | **Sub-zone** | **County** | **Residence** |
| RFZ | Semi-Arid 2 (SA2) | | Urban | 6 |
| Rural | 38 |
| Northern Arid Lands (NAL) | Samburu | Urban | 2 |
| Rural | 3 |
| Northern Arid Lands (NAL) | Non-Samburu | Urban | 11 |
| Rural | 28 |
| **Total (RFZ stratum)** | | | | **88** |
| Non-RFZ | High Rainfall 1 (HR1) | | Urban | 15 |
| Rural | 73 |
| **Total (non-RFZ stratum)** | | | | **88** |
| **Overall total** | | | | **176** |

Overall, the main sample is made up of 176 EAs, with 88 EAs selected for the RFZ sample (i.e., within the RFZ stratum) and 88 EAs selected within the HR1 counties not covered by the RFZ (i.e., within the non-RFZ stratum).

**Parallel survey EA sample**  
The parallel survey sample will be selected as an additional sample of 2 households in the same EAs as the main survey. Therefore, the first-stage sample of EAs is the same for the main and parallel surveys.

### 6.3.2 Preparation of the second- and third-stage sampling frames: Listing

Before conducting the second and subsequent stages of sampling, a field-based listing operation must take place within each sampled EA. The listing operation is critical because it provides an update of the current situation on the ground in these areas in terms of the sampling units and target population. This information is needed to create an up-to-date and accurate sampling frame for subsequent stages of sampling.

The 176 EAs selected for the main sample during the first stage of sampling are beginning to undergo the listing process. During the listing process, some of the EAs might be identified to undergo segmentation (which involves an additional stage of selection) based on the given criteria (see Section 6.3.3). For EAs that do not require segmentation, a listing of all households will be undertaken following the listing methodology. The listing methodology for the survey is based on the recommended listing guidelines for Feed the Future PBSs.

**Parallel survey listing**  
The parallel survey sample of households will be selected from the same listing of households used for the main survey sample. Therefore, the list frame of households is the same for the main and parallel surveys.

### 6.3.3 Potential second-stage sampling of segmented EAs

There could be substantial changes (i.e., growth or reduction) in the size of some EA populations between the last count and the time when the PBS will be conducted. If a sampled EA has grown too large by the time of the listing operation, field teams need to divide the EA and subsample one part of it, through a process called segmentation.

For the Midline Survey, the first-stage sampling frame is based on the 2019 census EAs within the P2-ZOI and RFZ. Given the time that has passed since the census, some changes on the ground in terms of the size of some of the EAs could be expected. Based on the average EA size from the 2019 census count, the criterion used to identify EAs that require segmentation is when their size exceeds 200 households, as recommended in the Feed the Future PBS sampling guide.

After an EA has been identified for segmentation, it is divided into an appropriate number of segments such that each segment is roughly equal to the average EA size with clearly identifiable boundaries, as recommended in the Feed the Future PBS sampling guide.[[17]](#footnote-16) One segment will randomly be selected from the EA using PPS sampling (with the MOS being the number of dwelling units identified per segment based on a rough count). The selected segment is then listed to create a list frame of households within the segment. The segmentation process will be undertaken as part of the listing operations, following the detailed step-by-step instructions provided in the listing manual. Therefore, a cluster in the survey could be an EA or a segment of an EA.

After the listing operation is complete, an up-to-date list frame of the households within each sampled cluster will be available to be used to select the households for the next stage of sampling.

### 6.3.4 Third-stage sampling of households within sampled EAs or segments

The third stage of sampling consists of randomly selecting the households in which to conduct interviews for the main and parallel surveys within each sampled cluster. The listing exercise provides an ordered list of households for each sampled cluster to serve as a household sampling frame.

The process of randomly selecting households will be undertaken using an equal probability variant of systematic sampling called fractional interval systematic sampling. Systematic sampling is the preferred method of selection because it spreads the sample of selected households throughout the cluster and, in so doing, captures more within-cluster variation than, for example, using a simple random sampling method. The implementation of the fractional interval systematic sampling of households at this stage will follow the recommendations provided in the Feed the Future PBS sampling guide. The sample taken will be 25 households per cluster for the main survey and 2 households per cluster for the parallel survey.[[18]](#footnote-17)

The household sample selection for the main and parallel surveys will be implemented using a two-phase sampling approach. This allows for both samples to be selected following the sampling guide recommendations but at the same time ensuring that the samples do not overlap. During the first phase, a sample of 27 households will be selected from each cluster. During the second phase, 2 households will be randomly selected from the first phase sample for the parallel survey, and the remaining 25 households will form the main survey household sample.

**Parallel survey household sample**  
The parallel survey sample will be selected as an additional sample of households in the same clusters as the main sample using a two-phase sampling approach. The sample take for the parallel survey will be 2 households per cluster to maintain the minimum required parallel sample size of at least 300 households per survey. These 2 households will be sampled in addition to the 25 households sampled for the main survey from each cluster using a two-phase sampling approach.

### 6.3.5 Fourth-stage sampling of individuals within sampled households

During the fieldwork and as part of the household interview, a roster of household members will be established for each sampled household by obtaining information from a responsible adult member of the household; this serves as a fourth-stage sampling frame of individuals for all sampled households. Demographic information, including the name, relationship to the primary adult male decision-maker (or primary adult female decision-maker if the household has no primary adult male decision-maker), age, sex, and any other information needed to establish eligibility to respond to one or more subsequent questionnaire modules, will be collected on all individuals included in the roster. For the agricultural technologies for the value chain commodities (VCCs) (i.e., millet, mango, dairy cows, and goats) modules, eligibility is determined in Module 2 (Dwelling characteristics and household assets) of the questionnaire, including age as reported in the roster as part of the eligibility criteria.

The Feed the Future PBS sampling guide recommends that all eligible individuals (for all target populations) residing within the sampled households should be selected for interview (i.e., eligible household members are sampled with certainty). This “take-all” approach is beneficial because it simplifies the sample design and fieldwork process. By including all eligible individuals with certainty, there is no need for separate individual-level weights to be derived at this stage that would increase the variability in survey weights and result in decreased precision of survey estimates for these target populations. Furthermore, selecting all eligible household members simplifies fieldwork because a procedure for selecting individuals within sampled households (such as the implementation of a Kish grid) does not need to be administered in the field. In addition, for more rare target populations, such as children 0-5 months of age, children 6-23 months of age, and children 0-59 months of age, this approach has the advantage of requiring the sampling of fewer households to achieve the desired number of sampled children. Finally, the approach reduces the design effect due to unequal weighting because all eligible members of the household are sampled with certainty.

**Table 8** lists the sampling groups at individual level that will be required for the 2022 Kenya Midline Survey.

**Table 8: Individual-level Sampling Groups**

| **Questionnaire module** | **Eligibility determined** | **Sampling group** | **Sampling approach** |
| --- | --- | --- | --- |
| Module 4: Women’s Nutrition (including program participation questions on women’s health and nutrition) | Roster + Module 4 | Women 15-49 years of age | Take all |
| Module 5: Children’s Nutrition  (including program participation questions on children’s health and nutrition) | Roster + Module 5 | Children under 5 years of age | Take all |
| Module 5: Children’s Nutrition  (including program participation questions on children’s health and nutrition)—EBF indicator | Roster + Module 5 | Children 0-5 months of age | Take all |
| Module 5: Children’s Nutrition  (including program participation questions on children’s health and nutrition)—Minimum acceptable diet indicator | Roster + Module 5 | Children 6-23 months of age | Take all |
| Module 6: Women’s Empowerment in Agriculture | Roster + Module 6 | Primary adult female decision-maker in the household  (age 18 or older) | Take all |
| Module 7.10: Agricultural Technologies—Millet | Roster (age) + Module 2  (responsible for cultivating millet in the past 12 months) | Household members 15+ years of age who have cultivated millet in the past 12 months and are responsible for making management decisions about the crops | Take all |
| Module 7.12: Agricultural Technologies—Mango | Roster (age) + Module 2  (responsible for cultivating mango in the past 12 months) | Household members 15+ years of age with at least 10 mango trees who have cultivated mango in the past 12 months and are responsible for making management decisions about the crops | Take all |
| Module 7.50: Agricultural Technologies—Dairy Cows | Roster (age) + Module 2  (responsible for raising dairy cows in the past 12 months) | Household members 15+ years of age who have raised dairy cows in the past 12 months and are responsible for making management decisions about the livestock | Take all |
| Module 7.51: Agricultural Technologies—Goats | Roster (age) + Module 2  (responsible for raising goats in the past 12 months) | Household members 15+ years of age who have raised goats in the past 12 months and are responsible for making management decisions about the livestock | Take all |

# Reserve sample

As mentioned previously, provision is made for the selection of a reserve sample of EAs to be used if some of the EAs from the main sample cannot be enumerated during fieldwork.

For the 2023 Kenya Midline Survey, the local survey implementation partner anticipates that some sampled EAs may require replacement using a reserve sample, given the security situation in certain areas covered by the P2-ZOI and RFZ. Based on discussions and the recommendations of the local survey implementation partner, counties in which there is a potential risk of security issues were identified. This was used as a basis for allocating the reserve sample across the strata for the survey.

The number of reserve EAs to be selected per stratum was based on the proportion of households in each stratum identified to be at-risk in terms of inaccessibility due to security concerns relative to the total number of households in that stratum.

**Table 9** presents the distribution of the reserve sample and the main sample EAs across the strata for the 2023 Kenya Midline Survey. In total, 26 reserve EAs were sampled, with most of the reserve allocated to the NAL sub-zone strata, because most of the potentially inaccessible areas were identified within these strata.

**Table 9: Distribution of the Main and Reserve Sample EAs per Stratum**

| **Stratum** | | | | **Main sample** | **Reserve sample** | **Percent of at-risk households per stratum** |
| --- | --- | --- | --- | --- | --- | --- |
| **Primary** | **Sub-zone** | **County** | **Residence** |
| RFZ | Semi-Arid 2 (SA2) | | Urban | 6 | 1 | 0% |
| Rural | 38 | 1 | 0% |
| Northern Arid Lands (NAL) | Samburu | Urban | 2 | 1 | 72% |
| Rural | 3 | 1 | 48% |
| Northern Arid Lands (NAL) | Non-Samburu | Urban | 11 | 5 | 43% |
| Rural | 28 | 15 | 54% |
| **Total (RFZ stratum)** | | | | **88** | **24** | **25%** |
| Non-RFZ | High Rainfall 1 (HR1) | | Urban | 15 | 1 | 0% |
| Rural | 73 | 1 | 0% |
| **Total (non-RFZ stratum)** | | | | **88** | **2** | 0% |
| **Overall total** | | | | **176** | **26** | **8%** |

Overall, for the RFZ, 25 percent of households are potentially at-risk particularly within the NAL sub-zone and these strata may require reserve sample EAs. Using these proportions for each design stratum, the number of EAs to be selected for the reserve EA sample was determined. Strata in which no at-risk households were identified such as the SA2 and HR1 sub-zone strata were still allocated one reserve sample EA in case of any unforeseen need.

It should be noted that the reserve sample should only be used as a last resort. The fieldwork teams should make every effort to enumerate in the 176 EAs that are part of the main sample before considering replacing any EA from the main sample.

The selection of the reserve sample was based on a two-phased sampling approach applied during the first-stage sampling of EAs (as described in Section 6.3.1), as recommended in the Feed the Future PBS sampling guide. During implementation of the survey, reserve sample EAs will be released based on the order of sampling, as recommended in the Feed the Future PBS sampling guide.

# Sample weighting

After data collection, processing, and editing, the final edited data from a PBS need to be weighted to account for the sample design and any adjustments due to non-response that may have occurred during data collection, before data analysis can take place. Sampling weights will be constructed based on the analysis requirements for the main and parallel surveys and to enable estimates to be produced for the overall P2-ZOI and RFZ. A community-level sampling weight will be constructed for analysis based on the community questionnaire. Household-level sampling weights will be constructed for the main and parallel surveys, taking into account the sample design for each survey. The main survey will also require individual-level sampling weights to be constructed for the required sub-populations. Individual weights for the following sub-populations will be constructed:

* Women 18 years of age or older who are the primary female decision-maker in the household
* Women 15-49 years of age
* Children under 6 months of age
* Children 6-23 months of age
* Children under 5 years of age
* Farmers 15 years of age or older, responsible for cultivating millet on at least one plot in the 12 months preceding the survey
* Farmers 15 years of age or older with at least 10 mango trees, responsible for cultivating mango on at least one plot in the 12 months preceding the survey
* Farmers 15 years of age or older, responsible for raising dairy cows in the 12 months preceding the survey
* Farmers 15 years of age or older, responsible for raising goats in the 12 months preceding the survey
* Farmers 15 years of age or older, responsible for cultivating one or more of the targeted VCC crops (i.e., millet or mango) on at least one plot or raising one or more herds of dairy cows or goats in the 12 months preceding the survey

This section provides an overview of the proposed methodology in the construction of these sampling weights, following the guidelines and recommendations in the Feed the Future PBS sampling guide.

## Calculating probabilities of selection and design weights

The sample is based on a stratified, multi-stage cluster sample design. At the first stage of selection, EAs are sampled as PSUs from the 2019 Kenya Population and Housing Census EAs that fell within the P2-ZOI and RFZ counties using a two-phase sampling approach to select a reserve sample of EAs in addition to the initial sample of EAs for the survey (see section 6.3.1), with the design weights being adjusted accordingly to account for this two-phase approach used during the sample design. The MOS used was the number of households based on the census count within these EAs.

Some sampled EAs may also need to undergo segmentation, with one segment being randomly selected for the survey. Therefore, the sampled clusters in the survey can be an EA or a segment of a sampled EA. For sampled clusters that are segments, the design weights will be adjusted accordingly to account for this additional stage of selection.

Within the selected clusters, a listing of households will be undertaken, and a sample of households will be selected from these listings using fractional interval systematic sampling. Within each sampled cluster, a fixed number of 25 households is expected to be selected for the main survey and 2 households for the parallel survey.

The sampling of households at this stage will be undertaken using a two-phase sampling approach to enable the selection of samples for both the main and parallel surveys, so the design weights for the parallel and main surveys will be adjusted to account for this two-phase approach.

Finally, within selected households, eligible individuals for each sampling group are considered for the survey. For the 2023 Kenya Midline Survey, all eligible individuals for a particular sampling group per household will be included for the survey (i.e., a take-all sampling approach). Therefore, the probability of selection for this stage of sampling will be one, as well as the design weight.

The overall selection probability will then be calculated as the product of the probabilities of selection across all stages of sampling, with the overall design weight taken as the inverse of the overall selection probability. Note that there will be a separate set of selection probabilities and design weights derived for the main and parallel surveys.

## Adjustments to the design weights

During execution of survey operations, issues may arise that affect the implementation of the proposed sample. It is important to try to address or at least account for these issues or changes as part of the sampling methodology because they may introduce bias or a decrease in the expected precision of the sample. These issues or changes can be addressed and accounted for through adjustments to the design weights. The following design weight adjustments are anticipated for the 2023 Kenya Midline Survey:

* **Weight adjustments for household non-response:** When implementing sample surveys in practice, it is usually the case that some proportion of households will be unreachable, unavailable, or unwilling to participate in the survey, despite the best efforts of field staff to enumerate these households. The Feed the Future PBS sampling guide recommends the use of a weighted household non-response rate in computing the non-response adjustment. This adjustment will be computed at the stratum level to preserve the self-weighting design. This adjustment will be implemented separately for the main and parallel survey samples.
* **Weight adjustments for individual non-response:** Individual non-response occurs when an eligible individual within a sampled household is not available for, refuses, or does not respond to a request for an interview. For individual non-response, a separate adjustment must be made for each sampling group to which the individual belongs, because non-respondents may be different for each sampling group. The recommended method to calculate the weight adjustments for individual non-response for any of the sampling groups in question is like the method used for calculating the household-level non-response adjustment factor but uses the weighted individual non-response rate in computing the non-response adjustment for a particular sampling group. This adjustment will be implemented separately for the main and parallel survey samples.

## Final sampling weights

The final sampling weights to be used during data analysis will be calculated differently, depending on whether the indicators to which they will be applied have underlying sampling groups at the individual level or at the household level (i.e., there will be separate weights that will be derived for household-level analysis and individual-level analysis).

For household-level indicators, the final sampling weights will be calculated by taking the product of design weights (which account for the selection probabilities during sampling) and the other weight adjustments (excluding individual-level weight adjustments). For the individual level, the final sampling weights will be the product of the final household sampling weights, together with other weight adjustments that have been calculated at the individual level.

**Table 10** lists the key indicators that will be directly collected or imputed from the survey and their respective sampling weights to be used during analysis.[[19]](#footnote-18),[[20]](#footnote-19)

**Table 10: List of Directly Collected or Imputed Indicators and Their Sampling Weights**

| **#** | **Indicator** | **Module** | **Survey weight** |
| --- | --- | --- | --- |
| 1 | Percentage of households with access to a basic sanitation service | Module 2: Dwelling Characteristics and Household Assets | Household |
| 2 | Percentage of households with soap and water at a handwashing station on premises | Module 2: Dwelling Characteristics and Household Assets | Household |
| 3 | Prevalence of water insecurity, based on the Brief Water Insecurity Experiences Scale (HWISE-4) | Module 2: Dwelling Characteristics and Household Assets | Household |
| 4 | Percentage of households below the comparative threshold for the poorest quintile of the Asset-Based Comparative Wealth Index | Module 2: Dwelling Characteristics and Household Assets | Household |
| 5 | Prevalence of moderate and severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES) | Module 3: Food Security | Household member\* |
| 6 | Ability to recover from shocks and stresses index | Module R: Resilience | Household |
| 7 | Index of social capital at the household level | Module R: Resilience | Household |
| 8 | Proportion of households that believe local government will respond effectively to future shocks and stresses | Module R: Resilience | Household |
| 9 | Proportion of households participating in group-based savings, micro-finance, or lending programs | Module R: Resilience | Household |
| 10 | Proportion of producers who have applied targeted improved management practices or technologies | * Module 7.10: Agricultural Technologies—Millet * Module 7.12: Agricultural Technologies—Mango * Module 7.50: Agricultural Technologies—Dairy cows * Module 7.51: Agricultural Technologies—Goats | * Millet farmers 15 years of age or older * Mango farmers 15 years of age or older * Dairy cow farmers 15 years of age or older * Goat farmers 15 years of age or older * Farmers of one or more VCCs (millet, mango, dairy cows, goats), 15 years of age or older |
| 11 | Percent of children 6-23 months of age receiving a minimum acceptable diet | Module 5: Children’s Nutrition | Children 6-23 months of age |
| 12 | Prevalence of exclusive breastfeeding of children under 6 months of age | Module 5: Children’s Nutrition | Children under 6 months of age |
| 13 | Prevalence of women of reproductive age consuming a diet of minimum diversity | Module 4: Women’s Nutrition | Women 15-49 years of age |
| 14 | Components of the A-WEAI[[21]](#footnote-20) | Module 6: Women’s Empowerment in Agriculture | Primary adult female decision-maker |
| 15 | Prevalence of poverty: Percent of people living on less than $1.90/day 2011 purchasing power parity | Imputed using Module 2 (Dwelling characteristics and household assets) information as predictors in a survey-to-survey imputation model[[22]](#footnote-21) | Household member\* |
| 16 | Depth of Poverty of the Poor: Mean shortfall of the poor relative to the $1.90/day 2011 PPP poverty line | Imputed using Module 2 (Dwelling characteristics and household assets) information as predictors in a survey-to-survey imputation model[[23]](#footnote-22) | Household member\* |
| 17 | Percent of people who are ‘near-poor,’ living on 100 percent to less than 125 percent of the $1.90 2011 PPP poverty line | Imputed using Module 2 (Dwelling characteristics and household assets) information as predictors in a survey-to-survey imputation model[[24]](#footnote-23) | Household member\* |
| 18 | Resilience Indicator:  Humanitarian Food Assistance Averted | Module R: Resilience | Household |
| 19 | Resilience Indicator:  Index of Shock Exposure | Module R: Resilience | Household |
| 20 | Resilience Indicator:  Absorptive capacity index | Module R: Resilience; Community Questionnaire | Household |
| 21 | Resilience Indicator:  Adaptive capacity index | Module R: Resilience; Community Questionnaire | Household |
| 22 | Resilience Indicator:  Transformative capacity index | Module R: Resilience; Community Questionnaire | Household |
| 23 | Resilience Indicator:  Livelihood Coping Strategies Index | Module 3: Food Security | Household |
| 24 | Resilience Indicator:  Index of Community Resilience | Module R: Resilience; Community Questionnaire | Household |
| 25 | Resilience Indicator:  Food Consumption Score | Module 3: Food Security | Household |
| 26 | Resilience Indicator:  Household Hunger Scale | Module 3: Food Security | Household |

In addition to the key indicators listed in **Table 10**, there are other survey results that will be presented in the midline indicator assessment report, covering various topics relating to climate adaptation, COVID-19, and program participation. **Table 11** lists these topics and their respective sampling weights to be used during analysis.

**Table 11: List of Additional Topics and Their Sampling Weights**

| **Topic** | **Module** | **Survey weight** |
| --- | --- | --- |
| Climate adaptation | Module 2A: Climate Adaptation | Household |
| COVID-19 | Module 3: Food Security | Household |
| Program participation, agriculture, and WASH | Module 2: Dwelling Characteristics and Household Assets | Household |
| Program participation, resilience | Module R: Resilience | Household |
| Program participation, women’s health, and nutrition | Module 4: Women’s Nutrition | Women 15-49 years of age |
| Program participation, children’s health, and nutrition | Module 5: Children’s Nutrition | Children under 5 years of age |

**Parallel survey sample weighting**  
The parallel survey household sample was selected as part of a two-phase sample design approach with the main survey household sample, so the household sampling weights for the parallel survey will be derived following a process like what is described for the main survey above, but with inputs at the household stage from the parallel sample. There are no individual-level sampling weights required for the parallel survey.

# Calculating measures of precision

The 2023 Kenya Midline Survey is a PBS that has a probability sample design and includes complex sample design features, such as unequal selection probabilities, stratification, and clustering. These sample design effects require specialized methodologies and procedures to correctly estimate measures of precision, such as standard errors, confidence intervals, and coefficients of variation, and to conduct valid statistical tests on survey estimates from these types of surveys. Most statistical software packages include functionality for analyzing data from complex sample surveys and correctly estimating measures of precision as well as running valid statistical tests. For most Feed the Future indicators, data analysis procedures are set up for use in Stata (except for the FIES indicator), and therefore the *Svy* commands should be implemented when analyzing data in Stata to make valid statistical inferences. The FIES indicator is calculated using R, therefore the *R Survey* package needs to be implemented to make valid statistical inferences for this indicator.

For more details on how to analyze complex sample survey data, please refer to Chapter 11 of the Feed the Future PBS sampling guide.

# References

*Guidance on the Feed the Future Phase Two Zone of Influence Midline Indicator Assessment.* Feed the Future Monitoring and Evaluation Guidance Series. Revised April 2022.

Stukel, D.M. (2018). *Feed the Future Population-Based Survey Sampling Guide.* Washington, DC: Food and Nutrition Technical Assistance Project, FHI 360.

Surveys for Monitoring for Resilience and Food Security (SMRFS). (2022*). Feed the Future Phase Two Zone of Influence Midline Indicator Assessment and Resilience Focus Zone Baseline Survey Scope of Work.* 2023 Kenya P2-ZOI Midline Indicator Assessment and RFZ Baseline Survey.

Surveys for Monitoring for Resilience and Food Security (SMRFS). (2023). *Zone of Influence Midline and Resilience Focus Zone Baseline Survey Protocol.* 2023 Kenya P2-ZOI Midline and RFZ Baseline Indicator Assessment. Rockville, MD: SMRFS.

# Annex: List of Sub-zones and counties comprising the Kenya phase two zone of influence and resilience focus zone

| **Sub-Zone** | ​​     ​**County** |
| --- | --- |
| High Rainfall 1 (HR1) | Bungoma County |
| Busia County |
| Homa Bay County |
| Kakamega County |
| Kisii County |
| Kisumu County |
| Migori County |
| Siaya County |
| Vihiga County |
| Semi-Arid 2 (SA2) | Kitui County |
| Makueni County |
| Taita Taveta County |
| Northern Arid Lands (NAL) | Garissa County |
| Isiolo County |
| Marsabit County |
| Turkana County |
| Wajir County |
| Samburu County\* |

\* Samburu county was a new county added to the P2-ZOI and RFZ in the NAL sub-zone in 2022 and therefore was not part of the counties covered at baseline.

1. Stukel, DM. 2018. Feed the Future Population-Based Survey Sampling Guide. Washington, DC: Food and Nutrition Technical Assistance Project, FHI 360. Available for download from <https://agrilinks.org/post/feed-future-zoi-survey-methods>. [↑](#footnote-ref-0)
2. Guidance on the Feed the Future Phase Two Zone of Influence Midline Indicator Assessment. Feed the Future Monitoring and Evaluation Guidance Series. Bureau for Resilience and Food Security. Revised April 2022. [↑](#footnote-ref-1)
3. Feed the Future Phase Two Zone of Influence Midline Indicator Assessment and Resilience Focus Zone Baseline Survey Scope of Work, Kenya 2023. September 2022. [↑](#footnote-ref-2)
4. The NAL sub-zone in the 2019 Kenya P2-ZOI Baseline Survey did not include Samburu county. The inclusion of Samburu county in the NAL sub-zone was finalized by USAID/KEA in early 2022 and represents a revision to the P2-ZOI and RFZ since the baseline. [↑](#footnote-ref-3)
5. Refer to Section 6.3. for more detail on the stages of sampling being implemented for the survey. [↑](#footnote-ref-4)
6. The key characteristics of a high-quality sampling frame are that it is comprehensive, complete, and up to date. Please refer to the Feed the Future PBS sampling guide for a detailed discussion of the importance of a high-quality sampling frame. [↑](#footnote-ref-5)
7. When the average EA size is small (i.e., less than 150 households per EA), the Feed the Future PBS sampling guide recommends segmenting any EA that is at least twice the size of the average EA in the survey area. In cases where the average EA size is large (i.e., at least 150 households per EA), then the sampling guide recommends segmenting EAs that exceed 300 households in size. [↑](#footnote-ref-6)
8. Information collected in Module 2 (Dwelling Characteristics and Households Assets) of the questionnaire on involvement in agriculture and/or livestock production is also used to establish the sample frame for VCC producers. [↑](#footnote-ref-7)
9. See the scope of work document for the guidance on determining the RFZ sample size and the document Guidance on the Feed the Future Phase Two Zone of Influence Midline Indicator Assessment for the guidance on determining the sample size for the P2-ZOI midline. [↑](#footnote-ref-8)
10. The indicator *Prevalence of water insecurity, based on the Brief Water Insecurity Experiences Scale (HWISE-4)* was not part of the sample size calculations because it is a new indicator that was added at midline. The indicator on minimum acceptable diet for children 6 -23 months of age, was also not part of the calculations since baseline estimates for this indicator was not available due to an error in the survey skip logic in the questionnaire which prevented this indicator from being calculated for the baseline. [↑](#footnote-ref-9)
11. For estimation of mean indicators, the term Pest1-Pest in formula (1) will be replaced by Xest2, which is the estimated variance for the distribution of indicator *X*. [↑](#footnote-ref-10)
12. No baseline estimate for the design effect of FIES was available, therefore the median value of 5 from the Gallup World Poll sponsored by the Food and Agriculture Organization of the United Nations was used, following the M&E guidance on Feed the Future Phase Two Zone of Influence Midline Indicator Assessments. [↑](#footnote-ref-11)
13. Note that for the *Prevalence of exclusive breastfeeding of children under six months of age indicator*, RFS guidance indicates using a formula to estimate the minimum required household sample size for the survey to include at least 70 children under 6 months of age (see equation [3]). The household sample size computed for this indicator will be compared against the largest final household sample size required across the other indicators to determine the final required household sample size. [↑](#footnote-ref-12)
14. This adjustment is determined based on the Stukel-Deitchler Inflator using the Kenya P2-ZOI baseline data on the average household size and proportion of the population making up the given sub-population for each indicator. Please refer to Appendix A of the Feed the Future Population-based Survey Sampling Guide for a more detailed discussion of this adjustment factor. [↑](#footnote-ref-13)
15. For individual-level indicators, the final required sample size computation will include an additional adjustment to determine the number of households to contact for the given individual sub-population. [↑](#footnote-ref-14)
16. A minimum of two EAs were selected in each stratum to maintain a minimum level of statistical representativeness and to be able to estimate measures of precision accounting for all strata in the sample. For more detail on the steps undertaken in implementing the systematic PPS sampling method, please refer to the Feed the Future PBS sampling guide. [↑](#footnote-ref-15)
17. When demarcating segments, it is important to be able to have clearly identifiable boundaries for each segment, such as a road or footpath, so that the field teams can easily identify the area covered by a segment during listing and data collection. [↑](#footnote-ref-16)
18. A sample take of 2 households per cluster for the parallel survey ensures that the minimum required sample size of at least 300 households for the parallel survey is maintained. [↑](#footnote-ref-17)
19. In addition, there are 4 anthropometry indicators (i.e., prevalence of stunted (HAZ < -2) children under 5 (0-59 months), prevalence of wasted (WHZ < -2) children under 5 (0-59 months), prevalence of underweight (BMI < 18.5) women of reproductive age, and prevalence of healthy weight (WHZ ≤ 2 and ≥-2) among children under 5 (0-59 months)) that will be constructed using secondary data analysis from the 2022 Kenya Demographic and Health Survey (KDHS) for this midline survey. These indicators will use the sampling weights from the 2022 KDHS for calculating these indicators and therefore are not listed in Table 11. [↑](#footnote-ref-18)
20. Note that there are no official indicators calculated with the children under 5 years of age sampling weight. This sampling weight, together with some of the other individual sampling weights, will be used for analysis of the program participation findings. [↑](#footnote-ref-19)
21. The survey will only collect data from the primary adult female decision-makers. The survey will be able to compute the five domains of empowerment sub-index, track progress in the proportion of women who are empowered, and measure changes to and progress in the key constraints to empowerment in agriculture for women in the ZOI but not the gender parity sub-index of the A-WEAI, and thus will not be able to compute the A-WEAI score. [↑](#footnote-ref-20)
22. The survey-to-survey imputation model is developed using per capita consumption expenditure from the P1-ZOI to predict per capita consumption expenditure for the P2-ZOI and impute the *prevalence of poverty, depth of poverty of the poor,* and *percent of people who are ‘near-poor’* indicators. The model for the midline will be further validated using data from the parallel survey for which per capita consumption expenditure will be collected. [↑](#footnote-ref-21)
23. Ibid. [↑](#footnote-ref-22)
24. Ibid. [↑](#footnote-ref-23)