

## SAMPLING FRAME

Based on the requirements of the Terms of Reference (ToR) and following discussion with the ESMAP – World Bank Team, a nationally representative core sample of 3,300 households (HHs) was selected proportionally from Kenya’s 47 counties combined with several oversampled groups. The oversampled groups were pulled from the marginal counties and specific World Bank intervention areas. The intervention areas initially included KEEP and KEMP focus areas and slum areas. However, KEMP areas were yet to be defined at the time of the study and were not included in the sample as initially planned. The core sample and over-sampling are shown in Figure 1.

**FIGURE A3.1: SAMPLING FRAME**

Basic MTF Sampling	Oversampling	Notes
47 Counties 3,300 HHs <sup>1</sup>	14 marginalized counties 1,100 HHs <sup>2</sup>	1) Originally enumeration areas belonging to KEEP and Nairobi slum areas were to be pulled out from 47 counties and treated as three separate strata. However, the target areas of these programs were not defined in terms of geographic administrative units, and they could not be readily matched to specific enumeration areas  2) 14 marginalized counties oversample was summed up with the core sample thus concurrent data collection and non-differentiating selection of enumeration areas. The marginalized counties include Garissa, Isiolo, Kilifi, Kwale, Lamu, Mandera, Marsabit, Narok, Samburu, Taita Taveta, Tana River, Turkana, Wajir and West Pokot.
	Nairobi – Slum Area: Intervention 600 HHs <sup>3</sup>	
	Nairobi – Slum Area: Non-Intervention 600 HHs <sup>3</sup>	

- 1) Originally enumeration areas belonging to KEEP and Nairobi slum areas were to be pulled out from 47 counties and treated as three separate strata. However, the target areas of these programs were not defined in terms of geographic administrative units, and they could not be readily matched to specific enumeration areas.
- 2) 14 marginalized counties oversample was summed up with the core sample thus concurrent data collection and non-differentiating selection of enumeration areas. The marginalized counties include Garissa, Isiolo, Kilifi, Kwale, Lamu, Mandera, Marsabit, Narok, Samburu, Taita Taveta, Tana River, Turkana, Wajir and West Pokot.

The survey was administered using 7 survey instruments (questionnaires): (i) households, (ii) institutional – schools, (iii) institutional – health centers, (iv) institutional – places of worship, (v) institutional – government offices, (vi) community focus group discussion and (vii) mini-grid developers. The questionnaires were drafted by the World Bank ESMAP team and revised by the

Consulting team based on extended discussion and feedback from pretesting exercises done in Kajiado and Narok towns. Table 1 summarizes the target number of interviews per survey instrument.

## **PRIMARY SAMPLING UNIT**

The team had originally planned to use PSUs defined by KNBS in the most recent National Sample Survey and Evaluation Program V (NASSEP V). After several rounds of discussions between KNBS, ESMAP, and the implementation team, the team decided to develop a separate sampling frame, which is described below. One disadvantage of the NASSEP V sampling frame was seen to be its limitation in

KENYA | Access to electricity and clean cooking in Kenya based on the Multitier Framework survey and data analysis adequately covering the KEEP and slum (GPOBA) sub-regions targeted by this survey, as the NASSEP V frame's enumeration areas are representative of statistics only at the national and county level.

To define and select enumeration areas, a publicly available gridded mapping of Kenya's population produced by NASA's Socioeconomic Data and Applications Center (SEDAC) was used. The SEDAC mapping estimates are consistent with Kenya's 2009 national censuses, and have been updated to match the 2015 Revision of UN World Population Prospects (CIESIN, 2016). This data is presented as a continuous raster surface with each pixel (1Km by 1Km) representing the population density of that area. Using these population estimates as a starting point, GIS experts from the National Autonomous University of Mexico (UNAM) worked to develop a spatial algorithm that defines discrete population enumeration areas bounded by Kenya's 7,149 sub-locations. Each enumeration area is continuous in space and contains roughly 200 households, which is similar in size to the PSUs defined by NASSEP V. Enumeration areas were then selected following the stratification plan described. The gridded population estimates, and population enumeration areas are shown in Figure 2: Gridded population based on UN estimations (left) and enumeration areas with roughly 200 HHs (right)

## **STRATIFICATION**

Stratified random sampling was used in household selection for the core sample and marginal counties. Stratification was by county, rural/urban residence and electrification status.

*County* - HH selection for the core and marginal county samples was proportional to each County's population based on the 2009 census.

*Rural/Urban* - Counties were divided equally between urban and rural areas, based on definitions in the 2009 census resulting in a 50-50 rural-urban split at the national level.

*Electrification status* - Once selected, rural and urban enumeration areas within each county were categorized by electrification status and the sample divided in half between electrified and non-

electrified enumeration areas. Electrification status is defined based on the location of KPLC step-down transformers. Detailed maps of transformer locations were provided by KPLC. 600m buffers were drawn around each transformer to simulate the distance that electrified households are typically located from transformers. Clusters that lie fully or partially within the 600m buffer were considered electrified (even if some of the enumeration area lay outside the 600m radius). Clusters that do not overlap with the buffer are considered unelectrified. Clusters were randomly selected from the categorized clusters with a bias given to electrified clusters. This provided for a balance between the electrified and non-electrified households as the possibility of having non-electrified households within electrified clusters was high as compared to the non-electrified clusters where 100% of the households in these clusters are not connected. In areas where the split between electrified and non-electrified clusters was uneven, then the category with fewer clusters was given priority where all the clusters in that category were picked with the remaining clusters selected from the other category.

Based on this stratification process, there were 4 possible characterization of the enumeration areas: Urban/electrified, Urban/unelectrified, Rural/electrified and rural unelectrified.

## **HOUSEHOLD SELECTION FOR THE CORE SAMPLE**

To select households, the team followed one of two options described below.

*Option 1) High resolution satellite imagery that allow dwellings within the enumeration area to be readily identified*

The team used Google Earth images to geo-reference, list, and randomly select 12 household structures from within the enumeration area. Where the selected structure turned out to something other than a household (e.g. commercial buildings), the enumerator selected the nearest household by distance. This approach was applied in areas with high population densities: urban areas and, sometimes, rural areas with an even and dense distribution of households. This method was unsuitable in dense slum areas, middle income urban areas with many multi-family dwellings (flats and apartments) and sparsely populated pastoral areas, where EAs occupied several hundred square kilometers and individual dwellings were difficult to identify from remotely sensed images.

*Option 2)*

Options 2 was the application of random walks and was applied in areas where Option 1 was not applicable. This method was viewed as being non-probabilistic and was applied for a greater majority of the interviews. In this approach, enumerators were trained to randomly select houses within their enumeration area ensuring that the respondent houses were spread across the entire area.

## **OVERSAMPLING**

### *Underserved counties*

14 underserved counties, listed in Figure 1, were oversampled by 1,100 HHs. This is in *addition to the HHs from these counties included in the core sample*. These underserved counties consist of mainly rural populations and therefore sample selection for oversampling did not follow the 1:1 urban-rural stratification applied to the core sample. Similarly, electrification is much lower than national average and a 1:1 ratio was also not possible, particularly in rural areas. The result of this oversampling on the distribution of enumeration areas was fewer or no rural electrified areas in a county and majority of the electrified EAs falling within urban centers and outskirts of major administrative towns.

### *KEEP and Nairobi slum areas (GPOBA areas and non-GPOBA areas)*

Two areas with current or soon-to-be implemented World Bank Supported energy access projects were oversampled. These were KEEP and GPOBA intervention areas. Discussions at proposal and inception stage also included KEMP intervention areas. These were, however, not included in the survey implementation due to lack of data on the specific target areas.

For KEEP, a listing of target areas was provided by KPLC. The listing included the size of the target population. This allowed for proportional distribution of respondents across 31 randomly selected target intervention areas. A total of 351 KEEP area surveys were carried out. Oversampling for GPOBA slum intervention and non-intervention areas focused on slums within Nairobi county. Respondents were randomly selected in both areas for a target of 1,200 households. The KEEP areas did not have clear geographic demarcations and even some of the local KPLC staff could not clearly distinguish the boundaries. Most of the local teams were not familiar with the term “KEEP” but understood these to mean the GPOBA areas.

## **SELECTION OF ENUMERATORS AND SUPERVISORS**

EED Advisory has developed an expansive database of over 130 enumerators spanning the various counties. This database consists of enumerators that EED-A has worked with directly on previous engagements, as well as enumerators from the KEFRI and Kenya Red Cross Society (KRCS) database and partner clients and organizations. Enumerator selection tapped into this list. KRCS vehicles and volunteers (who were recruited to work as enumerators) were used for high-risk counties including Mandera, Tana River, Garissa, Wajir and Nairobi’s slum areas. EED Advisory signed a memorandum of understanding with KRCS in this regard.

The following criteria was applied in enumerator selection. Enumerators had to:

- Be conversant with (preferably resident in) the area in which they were conducting the interviews. Understand and speak the local language in the area they would be conducting the interviews. <sup>[1]</sup><sub>SEP</sub>

- Be familiar with the political and administrative boundaries of the survey areas. [L] [SEP]
- Be fully available for the entire duration for which the interviews would be carried out. [L] [SEP]
- Provide a valid original Kenyan National Identification Card. [L] [SEP]
- Provide proof of graduation from a University recognized in Kenya. Diploma holders with substantial data collection experience were considered especially in the underserved counties. [L] [SEP]
- Provide their current telephone contact. [L] [SEP]
- Attend an enumerator training session before undertaking any survey interviews. [L] [SEP]
- Demonstrate ability to collect data using CAPI. [L] [SEP] Selected enumerators were trained and tested. The training program was over a duration of two days where enumerators were trained on the questionnaires including definition of terms, interviewing etiquette and best practices in asking questions. They were also trained on the survey tools including ODK and SW Maps (elaborated on later). Mock interviews and a pre-test exercise were integrated in the training program. Each enumerator was assigned a unique identity number and a supervisor. [L] [SEP] The enumerators were under the leadership of regional supervisors who reviewed submitted data and performed random checks to certify that the information collected was accurate. [L] [SEP]

## **REGIONAL SUPERVISORS** [L] [SEP]

Regional supervisors were drawn from EED-A and SEI staff and were responsible for training and coordinating survey teams, checking the work done by enumerators and leading introduction meetings with the relevant authorities before embarking on the data collection. They were the first contact point in case of any technical and logistical challenges faced by the data collection enumerators. The supervisors participated in a pre-survey training before being deployed to a region. They carried out a series of checks on every survey to ensure that the data collected was consistent and realistic. The supervisors provided the enumerators with reports of any errors observed in verified submitted data to ensure errors were not carried forward. Enumerators received feedback from a team of about 15 data verifiers also contracted by EED-A to check data uploaded on ODK over the first two days of field data collection in each county. [L] [SEP]

## **PRE-TESTING SURVEY TOOLS AND SUPERVISORS' TRAINING** [L] [SEP]

Pre-testing was done in Kajiado county and Narok county between 21/09/2016 – 22/09/2016 and 09/11/2016 and 10/11/2016 to assess the expected duration of an interview, respondents' comprehension of the questions, logical flow of the questionnaire and structure of each question,

appropriateness of the answer options, use of CAPI, sampling strategy assumptions among other issues. This process also helped detect problems with the questionnaire design, highlight sensitive questions, and identify redundant and ambiguous questions. The supervisors' training was done on the 16/11/2016 and 17/11/2016 in Karen, on the outskirts of Nairobi. A process map to standardize the data collection process was developed Computer Aided Personal Interview Open Data Kit In carrying out the tablet-based surveys, we used Open Data Kit (ODK). We developed the necessary scripts required on ODK built from the provided questionnaires. The platform provided both data collection and data analysis capabilities, automatically uploading the data onto a central server, facilitating GPS positioning of the survey point and inputting a timestamp. ODK had been pre-tested and had proven capable of delivering data to the standard and intention of the Client. It was uploaded on all the tablets prior to distribution to the enumerators and supervisors. The set-up software is cross-compatible on Windows, Mac and Linux platforms. Detailed training on using ODK for this survey was provided to the supervisors and enumerators during the pre-testing period. Once the data is entered, the team would carry out a detailed analysis and provide a full array of descriptive statistics for all indicators of interest. Descriptive statistics will be disaggregated to indicate regional trends and contrasts between urban and rural sub-populations, as well as other socio-demographic groups (e.g. female-headed households). As expected, certain sites did not have access to power to charge the tablets therefore the enumerators being dispatched to these areas were given portable power sources to ensure their tablets were always charged during the interviews.

### *SW Maps*

The randomly selected enumeration areas, county sub-locations, and randomly selected households per county were overlaid on SW Maps, a GIS android application that can collect, present, and share geographic information. The App enabled enumerators to navigate to the enumeration areas in real time and provide live maps of the enumeration area, sub-locations, and the randomly selected households. Information (electrification status, rural urban status, and administrative information) on each enumeration area was also visible on the App. SW maps can also record GPS tracks and measure distances.

### *GIS Database*

The survey questionnaires were set up to collect the GPS coordinates of all the interviews carried out. To do this, each tablet's inbuilt location services was switched on to automatically record the GPS coordinates onto the ODK questionnaire. The tablet's location services and SW Maps facilitated the movement of enumerators and supervisors in the field as they could locate and plan for transport based on distances and accessibility. All the surveys with the correct spatial information (County name, District, Division, Location, Sub-location and Cluster Unique Id/EA Id) can be presented in shapefile(.shp) and Google Earth (.kml/.kmz) format.

