Rwanda Energy Survey (2022) Sampling Strategy and Weighting

This document explains the methodology of the 2022 Rwanda energy survey, starting with the sampling strategy and then the weighting process.

Sampling Strategy

1. Sample design and selection

The energy survey included a household survey, a public institution survey, and a survey of refugee camps and adjacent host communities. For each survey, different sampling frames were used to select representative samples.

(1) Household Survey

The sampling frame for the household survey was the complete list of residential households from the *Rwanda Population and Housing Census Report 2012 (RPHC2012)* provided by the National Institute of Statistics. It excluded households located within 5km from the refugee camp, and it comprised information on Rwandan residential households grouped by Province, District, Sector, Cell, and Village.

(2) Public Institutions Survey

For the public institutions survey, given that the sampling frame is not available, all education and health centers located in the area where the household survey took place were interviewed. The sample size for the public institutions survey depended on the number of institutions found in villages selected for the household survey at the time of survey administration.

(3) Refugee Households and Host Community Surveys

The refugee household survey was conducted in the five refugee camps in Rwanda located in different areas. The sampling frame for this survey was the exhaustive list of households in the five refugee camps. Additionally, adjacent host communities, within 5km of the refugee camps, were surveyed.

2. Household Survey: Sample Size Calculation, Stratification, and Sample Selection

Sample Size Calculation

The sample size for the household survey was designed to obtain estimates with high precision for the main indicators targeted by the survey. Since the Rwanda Demographic and Health Survey (2019/2020) indicated that in Rwanda, 46% of households have access to electricity, the value was used in the calculation of the sample size and considered as the benchmark indicator value. The minimum sample size required for the household survey was calculated using the formula (1).

$$(1) \text{ Sample size: } n = \text{Deft}^2 \frac{(1/p - 1)}{a^2}$$

Where:

n: *The minimum sample size for the household survey*
p: *Reference indicator.*
α: *Relative Standard Error (RSE);*

> *Deft* : *Design effect as the coefficient of adjustment of the size when the sample is drawn at different levels*

As the indicator of interest (P) is known, the sample size was calculated using the prevalence rate of 46% to maximize the sample size, with a precision level of 2.6% and a design effect of 1.5. Therefore, using the information described above, the minimum sample size required for this survey was 4,014 respondents (households).

Stratification

Stratification for the household survey was done considering province, urban, and rural area as strata to increase the efficiency of the sample design. An equal sample allocation was applied to allocate households between urban and rural areas and later a square root allocation was applied to allocate the sample size between provinces

Sample Selection

In the first stage, 223 villages were selected using probability proportional to size (PPS), and in the second stage, after listing all residential households in the sampled villages, 18 households were selected using systematic random sampling methods (it was anticipated that 50% of electrified households and 50% of non-electrified households be sampled from each sampled village for interview).

3. Public Institutions Survey: Sample Size

The sample size for the public institutions' survey corresponded to all educational institutions and health centers found in the sampled villages. In other words, the final sample size for the public institution survey was determined based on the number of villages selected for the household survey.

4. Refugee and Host Community Surveys: Sample Size and Sample Selection

Sample Size

The sample size for the refugee camp and hosting communities was determined using the same formula used to calculate the minimum sample size for the household survey. The only difference is that since the indicator of interest (P) is unknown (proportion of households connected to the national grid), the prevalence rate was assumed to be 50% to maximize sample size and relative standard error to 3.6%. Therefore, the minimum sample size was 1,700 households.

The total population of refugee camps and host communities was 264,316 of which the population of refugee camps represented 10% and the population of hosting communities represents 90%. To design an effective sample size for each category, 40% of the sample was drawn from the refugee camp and 60% from hosting communities. In other words, 700 households were selected for interview in refugee camps and 1,000 households in hosting communities; this implies a relative standard error (RSE) of 5.7% and 4.7% respectively.

Sample Selection

For the refugee household survey, in the first stage, 39 segments from refugee camps were selected using probability proportion to size (PPS), and in the second stage, a systematic random sampling technique was applied to select 18 refugee households from the sampled segment.

For host communities survey, 56 villages were selected using probability proportional to size (PPS) and at the second stage, 18 households from each sampled village were selected using systematic random sampling technics.

Sample Weight Calculation

The weights were the inverse of different selection probabilities at different stages.

(1) Household survey targeting Rwandan households and host community

Step 1: Probability of selection of villages in each stratum
Step 2: Probability of selection of households in the sampled village

a. Probability of selecting villages

Before selecting villages, provinces, and areas (urban and rural) were combined to form strata and from each stratum, a sample of villages was selected using probability proportional to size (PPS) as specified in the sampling methodology note. The PPS formula used to select the sample villages is as follows:

$$P_{1i} = \frac{k * MOS_i}{\sum_h MOS_1}$$

Where:
P_{1i} =Probability of selecting a village in strata h
k =Number of villages to be sampled in strata h
MOS_i =Number of households listed in census 2012 in village i
$\sum_h MOS_1$ =Total number of households in strata h

b. Probability of selecting households in the sampled village

After updating the number of households in the sampled villages during the listing operation and grouping together households connected to the national grid (group A) and those off-grid (group B), ideally, the sample households (18) should be composed of 50% (9) of the households connected to the national grid and 50% (9) of the households off-grid/not connected. If the households listed in a group (A or B) are less than 9, all households were selected, and the remaining households to complete the targeted 18 households in the sampled village were selected in the other group (B or A).

The probability of selecting the households for interview from each group was computed as follows:

$$P_{2iA} = \frac{n_A}{MOSL_A}$$
$$P_{2iB} = \frac{n_B}{MOSL_B}$$

Where:
P_{2iA} =Probability of selecting households connected to national grid in sampled village i
n_A: sample of households connected to national grid
$MOSL_A$: Number of households listed connected to the national grid in the sampled village i
P_{2iB} = Probability of selecting households off grid in sampled village i
n_B: Sample of households off grid
$MOSL_B$: Number of households listed off grid in sampled village i

After computing first probability and second probability, the final probability was computed as the product of the two probabilities.

- o For households connected to national grid
  $$P_{TA} = P_{1i} * P_{2iA}$$

- o For households off grid
  $$P_{TB} = P_{1i} * P_{2iB}$$

c. Calculation of the household's weight

The weights for analysis were computed as the inverse of the total probability, as follows:

- o For households connected to national grid
  $$W_{iA} = \frac{1}{P_{Tcg}} = \frac{\sum_h MOS_1}{k * MOS_i} * \frac{MOSL_A}{n_A}$$

- o For households off grid
  $$W_{iB} = \frac{1}{P_{TB}} = \frac{\sum_h MOS_1}{k * MOS_i} * \frac{MOSL_B}{n_B}$$

Where:
> W_{iA}: Weight for households connected to national grid in sampled village i
> W_{iB}: Weight for households off grid in sampled village i

d. Adjustment of non-response

To compute the final weight of analysis, a factor to adjust the non-response was applied to the household's weight specifically for those villages where several sampled households were less than 18. To do that, using the list of households sampled, the household was identified weather was sampled as connected to national grid or off grid/not connected. Then, a factor of adjustment was computed as the ration of total households sampled divided to the number of households interviewed in the specific group (connected to national grid or off grid).

- o Final households' weight for households connected to national grid
  $$HH\_WT_{iA} = W_{iA} * \frac{n_A}{n_{AC}}$$

Where:
> HH_WT_{iA}: Weight for households connected to national grid in sampled village i
> n_{AC}: Number of households connected to the national grid that completed the interview

- o Final households' weight for households off grid
  $$HH\_WT_{iB} = W_{iB} * \frac{n_B}{n_{BC}}$$
  HH_WT_{iB}: Weight for households off grid in sampled village i
  n_{BC}: Number of households off grid that completed interview

(2) Refugee households survey

a. Probability of selecting households in the refugee camp

The probability of selecting households for interview from each group was computed as follows:

$$P_i = \frac{n}{MOSL}$$

Where:

P_i =Probability of selecting households in the refugee camp i

n: Number of sampled households

$MOSL$: Total number of households in the refugee camp i

b. Calculation of the refugee household's weight

The weight for the refugee household analysis was computed the inverse of the total probability as follows:

$$W_i = \frac{1}{P_i}$$

Where:

W_i: Weight for the household in the refugee camp i