

Kenya - COVID-19 Rapid Response Phone Survey with Households 2020-2022, Panel

Nistha Sinha

Report generated on: September 21, 2022

Visit our data catalog at: <https://microdata.worldbank.org/index.php>

Identification

SURVEY ID NUMBER

KEN_2020_COVIDRS_v07_M

TITLE

COVID-19 Rapid Response Phone Survey with Households 2020-2022, Panel

SUBTITLE

Waves 1-8

COUNTRY/ECONOMY

Name	Country code
Kenya	KEN

STUDY TYPE

1-2-3 Survey, phase 1 [hh/123-1]

SERIES INFORMATION

This dataset contains information from the eight waves of the COVID-19 RRPS with households in Kenya. The first five waves were conducted over a period of two months each, while wave 6 and 7 were conducted over a period of four months. Wave 8 was conducted over four weeks. Data collection was implemented between May 2020 and July 2022.

The survey was conducted as follows:

Wave 1: May 14 to July 7, 2020; 4,061 Kenyan households

Wave 2: July 16 to September 18, 2020; 4,492 Kenyan households

Wave 3: September 28 to December 2, 2020; 4,979 Kenyan households

Wave 4: January 15 to March 25, 2021; 4,892 Kenyan households

Wave 5: March 29 to June 13, 2021; 5,854 Kenyan households

Wave 6: July 14 to November 3, 2021; 5,765 Kenyan households

Wave 7: November 15, 2021, to March 31, 2022; 5,633 Kenyan households

Wave 8: May 31, 2022, to July 08, 2022; 4,550 Kenyan households

ABSTRACT

The World Bank in collaboration with the Kenya National Bureau of Statistics and the University of California, Berkeley are conducting the Kenya COVID-19 Rapid Response Phone Survey to track the socioeconomic impacts of the COVID-19 pandemic, the recovery from it as well as other shocks to provide timely data to inform policy. This dataset contains information from eight waves of the COVID-19 RRPS, which is part of a panel survey that targets Kenyan nationals and started in May 2020. The same households were interviewed every two months for five survey rounds, in the first year of data collection and every four months thereafter, with interviews conducted using Computer Assisted Telephone Interviewing (CATI) techniques.

The data set contains information from two samples of Kenyan households. The first sample is a randomly drawn subset of all households that were part of the 2015/16 Kenya Integrated Household Budget Survey (KIHBS) Computer-Assisted Personal Interviewing (CAPI) pilot and provided a phone number. The second was obtained through the Random Digit Dialing method, by which active phone numbers created from the 2020 Numbering Frame produced by the Kenya Communications Authority are randomly selected. The samples cover urban and rural areas and are designed to be representative of the population of Kenya using cell phones. Waves 1-7 of this survey include information on household background, service access, employment, food security, income loss, transfers, health, and COVID-19 knowledge and vaccinations. Wave 8 focused on how households were exposed to shocks, in particular adverse weather shocks and the increase in the price of food and fuel, but also included parts of the previous modules on household background, service access, employment, food security, income loss, and subjective wellbeing.

The data is uploaded in three files. The first is the hh file, which contains household level information. The 'hhid', uniquely identifies all household. The second is the adult level file, which contains data at the level of adult household members. Each adult in a household is uniquely identified by the 'adult_id'. The third file is the child level file, available only for waves 3-7, which contains information for every child in the household. Each child in a household is uniquely identified by the 'child_id'.

The duration of data collection and sample size for each completed wave was:

Wave 1: May 14 to July 7, 2020; 4,061 Kenyan households

Wave 2: July 16 to September 18, 2020; 4,492 Kenyan households
 Wave 3: September 28 to December 2, 2020; 4,979 Kenyan households
 Wave 4: January 15 to March 25, 2021; 4,892 Kenyan households
 Wave 5: March 29 to June 13, 2021; 5,854 Kenyan households
 Wave 6: July 14 to November 3, 2021; 5,765 Kenyan households
 Wave 7: November 15, 2021, to March 31, 2022; 5,633 Kenyan households
 Wave 8: May 31 to July 8, 2022; 4,550 Kenyan households

The same questionnaire is also administered to refugees in Kenya, with the data available in the UNHCR microdata library:
<https://microdata.unhcr.org/index.php/catalog/296/>

UNIT OF ANALYSIS

Household, Individual

Version

VERSION DESCRIPTION

Version 07.

Changes made since last update:

- 1) Wave 8 data was added;
- 2) Added 3 observations to wave 7 data which were previously incorrectly dropped

VERSION DATE

2022-08-22

VERSION NOTES

Changes made since last update:

- 1) Wave 8 data was added;
- 2) Added 3 observations to wave 7 data which were previously incorrectly dropped

Scope

NOTES

The Kenya COVID-19 RRPS survey covers the following topics: Household Roster, Travel Patterns & Interactions, Employment, Food security, Income Loss, Transfers, Subjective welfare (50% of sample), Health, COVID-19 Knowledge and Vaccinations, Household and Social Relations (50% of sample). In wave 8, the questionnaire was shortened: modules on Health, COVID-19 Knowledge and Vaccinations were dropped and only essential questions were kept in the remaining modules. New questions were added on the exposure to idiosyncratic and aggregate shocks, on food and fuel price increases and subjective wellbeing.

Coverage

GEOGRAPHIC COVERAGE

National coverage covering rural and urban areas

Producers and sponsors

PRIMARY INVESTIGATORS

Name

Nistha Sinha

PRODUCERS

Name

University of California Berkeley
Kenya National Bureau of Statistics

FUNDING AGENCY/SPONSOR

Name

IBRD

Sampling

SAMPLING PROCEDURE

The COVID-19 RRPS with Kenyan households has two samples. The first sample consists of households that were part of the 2015/16 KIHBS CAPI pilot and provided a phone number. The 2015/16 KIHBS CAPI pilot is representative at the national level stratified by county and place of residence (urban and rural areas). At least one valid phone number was obtained for 9,007 households and all of them were included in the COVID-19 RRPS sample. The target respondent was the primary male or female household member from the 2015/16 KIHBS CAPI pilot. The second sample consists of households selected using the Random Digit Dialing method. A list of random mobile phone numbers was created using a random number generator from the 2020 Numbering Frame produced by the Kenya Communications Authority. The initial sampling frame therefore consisted of 92,999,970 randomly ordered phone numbers assigned to three networks: Safaricom, Airtel and Telkom. An introductory text message was sent to 5,000 randomly selected numbers to determine if numbers were in operation. Out of these, 4,075 were found to be active and formed the final sampling frame. There was no stratification and individuals that were called were asked about the households they live in. Until wave 7 sampled households that were not reached in earlier waves were also contacted along with households that were interviewed before. In wave 8 only households that had previously participated in the survey were contacted for interview. The “wave” variable represents in which wave the households were interviewed in.

WEIGHTING

Cross-Sectional weights

For the KNBS and RDD samples, to make the sample nationally representative of the current population of households with mobile phone access, we create weights in two steps.

Step 1: Construct raw weights combining the two national samples: The current population consists of

- (I) households that existed in 2015/16, and did not change phone numbers,
- (II) households that existed in 2015/16, but changed phone number,
- (III) households that did not exist in 2015/16.

Abstracting from differential attrition, the weights from the 2015/16 KIHBS CAPI pilot make the KIHBS sample representative of type (I) households. For RDD households, we ask whether they existed in 2015/16, when they had acquired their phone number, and where they lived in 2015/16, allowing us to classify them into type (I), (II) and (III) households and assign them to KIHBS strata. We adjust weights of each RDD household to be inversely proportional to the number of mobile phone numbers used by the household, and scale them relative to the average number of mobile phone numbers used in the KIHBS within each stratum. RDD therefore gives us a representative sample of type (II) and (III) households. We then combine RDD and KIHBS type (I) households by ex-post adding RDD households into the 2015/16 sampling frame and adjusting weights accordingly. Last, we combine our representative samples of type (I), type (II) and type (III), using the share of each type within each stratum from RDD (inversely weighted by number of mobile phone numbers). Variable: `weight_raw`

Step 2: Scale the weights to population proportions in each county and urban/rural stratum: We use post stratification to adjust for differential attrition and response rates across counties and rural/urban strata. We scale the raw weights from step 1 to reflect the population size in each county and rural/urban stratum as recorded in the 2019 Kenya Population and Housing Census conducted by the KNBS (2019 Kenya Population and Housing Census, Volume II: Distribution of Population by Administrative Units, December 2019, Kenya National Bureau of Statistics, <https://www.knbs.or.ke/?wpdmpo=2019-kenya-population-and-housing-census-volume-ii-distribution-of-population-by-administrative-units>). Variable: `weight`

In addition to being nationally representative, the data is also weekly representative for all waves except for wave 8. The variable `weight_weekly` should be used for weekly representative estimates.

Panel Weights

To construct panel weights, we follow the approach outlined in Himelein (2014): “Weight Calculations for Panel Surveys with Subsampling and Split-off Tracking”. In each household we follow one target respondent. Wherever households split, only the current household of the target respondent was interviewed. The weights for the wave 1 and 2 balanced panel are constructed by applying the following steps to the full sample of Kenyan nationals:

0. Wave 1 cross-sectional weights after post-stratification adjustment are used as a base. $W_1 = W_{\text{wave1}}$
1. Attrition adjustment through propensity score-based method: The predicted probability that a sample household was successfully re-interviewed in the second survey wave is estimated through a propensity score estimation. The propensity score (PS) is modeled with a linear logistic model at the level of the household. The dependent variable is a dummy indicating whether a household that has completed the survey in wave 1 has also done so in wave. The following covariates were used in the linear logistic model: Urban/rural dummy, County dummies, Household head gender, Household head age, Household size, Dependency ratio, Dummy: Is anyone in the household working, Asset ownership: Radio, Asset ownership: Mattress, Asset ownership: Charcoal Jiko, Asset ownership: Fridge, Wall material: 3 dummies, Floor materials: 3 dummies, Connection to electricity grid, Number of mobile phones numbers household uses, Number of phone numbers recorded for follow-up, Sample dummy for estimation with national samples
2. Rank households by PS and split into 10 equal groups
3. Calculate attrition adjustment factor: ac (attrition correction) = the reciprocal of the mean empirical response rate for the propensity score decile
4. Adjust base weights for attrition: $W_2 = W_1 * ac$
5. Trim top 1 percent of the weights distribution (), by replacing the weights among the top 1 percent of the distribution with the highest value of a weight below the cutoff. $W_3 = \text{trim}(W_2)$
6. Apply post-stratification in the same way as for cross-sectional weights (step 2) Variable: `weight_panel_w1_2`

The balanced panel weights including waves 3, 4, 5, 6, 7 and 8 were constructed using the same procedure. Variables: `weight_panel_w1_2_3`, `weight_panel_w1_2_3_4`, `weight_panel_w1_2_3_4_5`, `weight_panel_w1_2_3_4_5_6`, `weight_panel_w1_2_3_4_5_6_7` and `weight_panel_w1_2_3_4_5_6_7_8`.

Data Collection

DATES OF DATA COLLECTION

Start	End	Cycle
2020-05-14	2020-07-31	Wave 1
2020-07-16	2020-09-18	Wave 2
2020-09-28	2022-12-02	Wave 3
2021-01-15	2021-03-25	Wave 4
2021-03-29	2021-06-13	Wave 5
2021-07-14	2021-11-03	Wave 6
2021-11-15	2022-03-31	Wave 7
2022-05-31	2022-07-08	Wave 8

DATA COLLECTION MODE

Computer Assisted Personal Interview [capi]

DATA COLLECTION NOTES

PRE-LOADED INFORMATION: Basic household information was pre-loaded in the CATI assignments for each enumerator. The information, for example the household's location, household head name, phone numbers etc, was used to help enumerators call and identify the target households. The list of individuals from the KIHBS CAPI pilot and their basic characteristics were uploaded as well as basic information from previous survey waves where available from wave 2 onward.

RESPONDENTS: The COVID-19 RRPS had one respondent per household. For the sample from the 2015/16 KIHBS CAPI pilot, the target respondent was defined as the primary male or female adult household member. They were randomly chosen where both existed to maintain gender balance. If the target respondent was not available for a call, the field team spoke to any adult currently living in the household of the target respondent. If the target respondent was deceased, the field team spoke to any adults that lived with the target respondent in 2015/16. Finally, if the household from 2015/16 split up, we targeted anyone in the household of the target respondent but did not survey a household member that no longer lives with

the target respondent. For the sample based on Random Digit Dialing, the target respondent was the owner the phone number that was randomly selected. Where the target respondent was not available for the interview, we spoke to any other adult household member of the target respondent.

Questionnaires

QUESTIONNAIRES

The questionnaire was administered in English and is provided as a resource in pdf format.

Additionally, questionnaires for each wave are also provided in Excel format coded for SCTO.

The same questionnaire is also administered to refugees in Kenya, with the data available in the UNHCR microdata library: <https://microdata.unhcr.org/index.php/catalog/296/>

Access policy

CONFIDENTIALITY

Before being granted access to the dataset, all users have to formally agree: 1. To make no copies of any files or portions of files to which s/he is granted access except those authorized by the data depositor. 2. Not to use any technique in an attempt to learn the identity of any person, establishment, or sampling unit not identified on public use data files. 3. To hold in strictest confidence the identification of any establishment or individual that may be inadvertently revealed in any documents or discussion, or analysis. Such inadvertent identification revealed in her/his analysis will be immediately brought to the attention of the data depositor.

CITATION REQUIREMENTS

Use of the dataset must be acknowledged using a citation which would include:

- the Identification of the Primary Investigator
- the title of the survey (including country, acronym and year of implementation)
- the survey reference number
- the source and date of download

Example:

World Bank, University of California Berkeley, Kenya National Bureau of Statistics (2021). Kenya - COVID-19 Rapid Response Phone Survey with Households 2020-2022, Panel, Waves 1-8 (COVIDRS).Ref: KEN_2020_COVIDRS_v07_M. Downloaded from [uri] on [date].

Disclaimer and copyrights

DISCLAIMER

The user of the data acknowledges that the original collector of the data, the authorized distributor of the data, and the relevant funding agency bear no responsibility for use of the data or for interpretations or inferences based upon such uses.

Metadata production

DDI DOCUMENT ID

DDI_KEN_2020_COVIDRS_v07_M_WB

PRODUCERS

Name	Abbreviation	Affiliation	Role
Development Data Group	DECDG	World Bank	Documentation of the study

DATE OF METADATA PRODUCTION

2022-09-21

DDI DOCUMENT VERSION

Version 07 (2022-09-21)

Changes made since last update:

- 1) Wave 8 data was added;
- 2) Added 3 observations to wave 7 data which were previously incorrectly dropped'

Data Dictionary

Data file	Cases	Variables
hh The hh file contains household level information. The 'hhid', uniquely identifies all households.	40235	2410
hh_adult The adult level file contains data at the level of adult household members. Each adult in a household is uniquely identified by the 'adult_id'.	96878	346
hh_child The child level file contains information for every child in the household. Each child in a household is uniquely identified by the 'child_id'.	39690	220