Kenya - Socio-economic impact of COVID-19 on refugees - Panel Study

UNHCR

Report generated on: November 17, 2022

Visit our data catalog at: https://microdata.unhcr.org/index.php

Overview

Identification

ID NUMBER
UNHCR_KEN_2020_COVID_Panel_v2.1

Version

VERSION DESCRIPTION
v2.1: Edited, anonymous dataset for licensed distribution. Seventh wave.

PRODUCTION DATE
2022-09

NOTES
Version Notes: Version 07. Changes made since last update

- Wave 8 data was added

- Added 3 observations to wave 7 data which were previously incorrectly dropped

Overview

ABSTRACT
The World Bank and UNHCR 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 a targeted response. This dataset contains information from eight waves of the COVID-19 RRPS, which is part of a panel survey that targets refugee household 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 sample aims to be representative of the refugee and stateless population in Kenya. It comprises five strata: Kakuma refugee camp, Kalobeyei settlement, Dadaab refugee camp, urban refugees, and Shona stateless. Waves 1-7 of this survey include information on household background, service access, employment, food security, income loss, transfers, health, and COVID-19 knowledge. 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; 1,328 refugee households

Wave 2: July 16 to September 18, 2020; 1,699 refugee households

Wave 3: September 28 to December 2, 2020; 1,487 refugee households

Wave 4: January 15 to March 25, 2021; 1,376 refugee households

Wave 5: March 29 to June 13, 2021; 1,562 refugee households

Wave 6: July 14 to November 3, 2021; 1,407 refugee households

Wave 7: November 15, 2021, to March 31, 2022; 1,281 refugee households

Wave 8: May 31 to July 8, 2022: 1,355 refugee households

The same questionnaire is also administered to nationals in Kenya, with the data available in the WB microdata library: <https://microdata.worldbank.org/index.php/catalog/3774>

KIND OF DATA
Sample survey data [ssd]

UNITS OF ANALYSIS
Individual and Household

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, Intentions/Solutions, and Household and Social Relations (50% of sample). In wave 8, the questionnaire was strongly adjusted: 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.

TOPICS

Topic	Vocabulary	URI
Protection		
Livelihood & Social cohesion		
Health		
Food Distribution		
Income Generation		

KEYWORDS
COVID-19, Refugees, Kenya, Economic monitoring

Coverage

GEOGRAPHIC COVERAGE
National coverage covering rural and urban areas

UNIVERSE
All persons of concern for UNHCR

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
UNHCR	

OTHER PRODUCER(S)

Name	Affiliation	Role
World Bank		
Kenya National Bureau of Statistics		
University of California		

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
UNHCR			

DATE OF METADATA PRODUCTION
2020-11-25

DDI DOCUMENT VERSION
1.0, November 2020

2.0, December 2020

3.0, February 2021

4.0, May 2021

5.0, July 2021

6.0, January 2022

7.0, July 2022

8.0, September 2022

DDI DOCUMENT ID
DDI_UNHCR_KEN_2020_COVID_Panel_v1.0

Sampling

Sampling Procedure

The sample aims to be representative of the refugee and stateless population in Kenya. It comprises five strata: Kakuma refugee camp, Kalobeyei settlement, Dadaab refugee camp, urban refugees, and Shona stateless, where sampling approaches differ across strata. For refugees in Kakuma and Kalobeyei, as well as for stateless people, recently conducted Socioeconomic Surveys (SES), were used as sampling frames. For the refugee population living in urban areas and the Dadaab camp, no such household survey data existed, and sampling frames were based on UNHCR's registration records (proGres), which include phone numbers. For Kakuma, Kalobeyei, Dadaab and urban refugees, a two-step sampling process was used. First, 1,000 individuals from each stratum were selected from the corresponding sampling frames. Each of these individuals received a text message to confirm that the registered phone was still active. In the second stage, implicitly stratifying by sex and age, the verified phone number lists were used to select the sample. 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. For the stateless population, all the participants of the Shona socioeconomic survey (n=400) were included in the RRPS, because of limited sample size. The sampling frames for the refugee and Shona stateless communities are thus representative of households with active phone numbers registered with UNHCR.

Weighting

Sampling weights for the refugee and stateless samples were tailored to the respective sampling strategies. Kakuma and Kalobeyei sub-samples have used the baseline weights from the respective SES underlying the sampling frame to adjust for any differences in the sampling probabilities. Then, propensity score weighting based on the full population covered in the SES household survey, have been used to account for differences in the probability of owning a phone number. The estimated propensity score reflects the household probability to have a phone number registered by UNHCR. To mitigate the effect of outlier estimates, the mean propensity score is computed for each decile. The baseline weights are then multiplied with the inverse of the propensity score deciles. For the refugees living in Dadaab camp and urban areas, a cell weighting approach has been used. Thereby, the sample is split into sub-groups (cells) based on the gender and age group of household head. The weights were then scaled such that they reflect the proportion of each cell in the UNHCR registration data of all refugees living in the respective location. In the group of stateless people registered with UNHCR, each household has the same weight assigned, as their full population is called in this survey. Lastly, to ensure sampling weights have the correct proportions across strata, they have been scaled to match population totals as provided by the up to date UNHCR registration data. Variable: weight

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. $W\_1 = W\_{wave1}$

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 = 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_8.

Questionnaires

Overview

The questionnaire included 12 sections
Section 1: Introduction
Section 2: Household background
Section 3: Travel patterns and interactions
Section 4: Employment
Section 5: Food security
Section 6: Income Loss
Section 7: Transfers
Section 8: Subjective welfare (50% of sample)
Section 9: Health
Section 10: COVID Knowledge
Section 11: Household and Social Relations (50% of sample)
Section 12: Conclusion

Data Collection

Data Collection Dates

Start	End	Cycle
2020-05-14	2020-07-07	1
2020-07-16	2020-09-18	2
2020-09-28	2020-11-30	3
2021-01-15	2021-03-25	4
2021-03-29	2021-06-13	5
2021-07-14	2021-11-03	6
2021-11-15	2022-03-31	7
2022-05-31	2022-07-08	8

Data Collection Mode

Computer Assisted Telephone Interview [cati]

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 UNHCR registration data and their basic characteristics were uploaded as well as basic information from previous survey waves where available from wave 2 onward.

Questionnaires

The questionnaire included 12 sections
Section 1: Introduction
Section 2: Household background
Section 3: Travel patterns and interactions
Section 4: Employment
Section 5: Food security
Section 6: Income Loss
Section 7: Transfers
Section 8: Subjective welfare (50% of sample)
Section 9: Health
Section 10: COVID Knowledge
Section 11: Household and Social Relations (50% of sample)
Section 12: Conclusion

Data Collectors

Name	Abbreviation	Affiliation
Vyxer		

Data Processing

Data Editing

Variable names were kept constant across survey waves. For questions that remained exactly the same across survey waves, data points for all waves can be found under one variable name. For questions where the phrasing changed (even in a minimal way) across waves, variable names were also changed to reflect the change in phrasing.

Extended missing values are used to indicate why a value is missing for all variables. The following extended missing values are used in the dataset:

· .a for 'Don't know'
· .b for 'Refused to respond'
· .c for 'Outliers set to missing'
· .d for 'Inconsistency set to missing' (used for employment data as explained below)
· .e for 'Field Skipped' (where an error in the survey tool caused the question to be missed)
· .z for 'Not administered' (as the variable was not relevant to the observation)

More detailed data on children was collected between waves 3 and 7, compared to waves 1, 2 and 8. In waves 1 and 2, data on children, e.g. on their learning activities, was collected for all children in a household with one question. Therefore, variables related to children are part of the 'hh' data for waves 1 and 2. Between waves 3 and 7, questions on children in the household were asked for specific children. Some questions covered all children, while others were only administered to one randomly selected child in the household. This approach allows to disaggregate data at the level of the child household members, and the data can be found in the 'child' data set. The household level weights can be used for analysis of the children's data. In wave 8, detailed information on children was dropped, as the questionnaire focused on other topics.

The education status of household members, except for the respondent, was imputed for rounds 1 and 2. For rounds 1 and 2, only the education status of the respondent was elicited, while for later rounds the education status for each household member was asked. In order to evaluate outcomes by the household member's education status, information on education was imputed for waves 1 and 2, using the information provided for all household members in waves 3, 4, and 5. This resulted in additional information on the education status for household members in round 1 and 2, which was not yet available for earlier versions of this data.

Some questions are not asked repeatedly across waves such that their values were imputed. For some questions, answers are not possible or unlikely to change within two months between survey waves such that households were not asked about them in all waves. The questions on assets owned before March 2020 were only asked to households when they are interviewed for the first time. The questions on the dwelling's wall and floor material as well as the household's connection to the power grid was not asked for all households in wave 2 and 3, where only new households and those who moved were covered by these questions. Questions on the main source of electricity in the households and types of assets owned were not asked in wave 8. The missing values those variables have when they were not asked, are imputed from the answers given in earlier waves.

Improved quality insurance algorithms lead to minor revisions to wave 1 to 5 data. Based on additional data checks, the team has made minor refinements to wave 1 to 5 data. The identification of the household members that were the respondent or the household head was refined in the rare cases where it was not possible to interview the same respondent as in previous waves for a given household such that another adult was interviewed. For this reason, for about 2 percent of observations the household head status was assigned to an incorrect household member, which was corrected. For <1 percent of households the respondent did not appear in adult level dataset. For about 1 percent of observations in wave 5 the respondent appeared twice in the adult level dataset.

Data from questions on COVID-19 vaccinations from wave 7 was dropped from the dataset. Due to significantly higher self-reported vaccination rates compared to official administrative records, data on vaccinations was deemed unreliable, most likely due to social desirability bias. Consequently, questions on vaccination status and questions using the vaccination data as a validation criterion were dropped from the datasets.

Data Appraisal

No content available

Documentation

Questionnaires

Questionnaire Wave 3

Title   Questionnaire Wave 3
Filename questionnaire_third_wave.pdf

Questionnaire Wave 1

Title   Questionnaire Wave 1
Filename kenya_covid_19_rrps_households_wave_1_questionnaire.pdf

Questionnaire Wave 4

Title   Questionnaire Wave 4
Filename Questionnaire-Fourth-Wave.pdf

Questionnaire Wave 6

Title   Questionnaire Wave 6
Filename Questionnaire-6th wave.pdf

Questionnaire Wave 5

Title   Questionnaire Wave 5
Filename Questionnaire-5th wave.pdf

Questionnaire Wave 2

Title   Questionnaire Wave 2
Filename HHQuestionnaire-Second-Wave.pdf

Questionnaire Wave 3

Title   Questionnaire Wave 3
Filename Questionnaire-3rd wave.pdf

Questionnaire Wave 7

Title   Questionnaire Wave 7
Filename kenya_covid_19_rrps_households_wave_7_questionnaire.pdf

Questionnaire Wave 8

Title   Questionnaire Wave 8
Filename   kenya_covid_19_rrps_households_wave_8_questionnaire.pdf

Reports

Report Wave 3

Title   Report Wave 3
Filename   Wave 3 Analysis.pdf

Report Wave 5

Title   Report Wave 5
Filename   Wave 5 Analysis.pdf

Report Wave 1

Title   Report Wave 1
Filename   Wave 1 Report.pdf

Report Wave 4

Title   Report Wave 4
Filename   Wave 4 Analysis.pdf

Technical documents

DDI

Title   DDI
Filename   ddi-documentation-english-296.pdf

Other materials

Monitoring COVID-19 Impact on Households in Kenya

Title   Monitoring COVID-19 Impact on Households in Kenya
Filename   https://www.worldbank.org/en/country/kenya/brief/monitoring-covid-19-impact-on-households-and-firms-in-kenya