



USAID
FROM THE AMERICAN PEOPLE

Feed the Future FEEDBACK
Northern Kenya 2015 Zones of Influence (ZOI) Interim Survey
Public Release Datasets and Documentation
README Data Description Document
January 2017

This deliverable package contains CSV, SAS, SPSS, and STATA datasets; a codebook; and a Privacy Assurance Statement (PAS) for the Northern Kenya 2015 ZOI Interim Survey.

Note that this survey included both the five counties (Garissa, Isiolo, Marsabit, Turkana, and Wajir) of the northern Kenya ZOI as well as the four comparison counties (Baringo, Mandera, Samburu, and Tana River) used for the Kenya REGAL impact evaluation (IE) analysis. In other words, quantitative data collected by this survey served the dual purpose of providing indicators for the northern Kenya Interim assessment as well as for the Kenya REGAL IE. The “zoi” flag variable in the dataset indicates whether a record is in the ZOI or in the comparison area. The “intensity” variable indicates whether a record is in a “high,” “medium,” or “low” programming intensity area.

Seventeen datasets are included with this deliverable: Household, Household Roster, Women, Children, three WEAL files, eight Expenditure files, and two Resilience files. (Module F in the Northern Kenya ZOI Interim Survey included the Resilience questions.) Note that all modules were collected in the Northern Kenya 2015 ZOI interim survey (i.e., all standard Feed the Future indicators for the northern Kenya interim assessment were calculated from primary ZOI interim survey data).

The Northern Kenya 2015 ZOI Interim Survey datasets contain observations from households with a completed interview. The variable order is generally the same as the questionnaire. New/derived variables (including intermediate and indicator variables), are labeled and typically located at the end of each file. (However, there is a separate “WEAL_Recode” dataset containing the derived variables from the WEAL analysis.)

Privacy Notice

These datasets are freely available to the public without restriction. However, all personal identification information that was present in the original data was removed. In addition, variables that could potentially be used indirectly (in combination with other variables or with other data sources) to identify any particular household or respondent were recoded to prevent that disclosure.

Location variables, unless otherwise noted, were randomized such that the location codes no longer reflect the original administrative codes or codes used in the sample frame. The anonymous codes are provided such that analysts may identify which records were collected from the same areas. The anonymous codes cannot be used to merge or compare with any other data source.

Households with large numbers of members can potentially be identified if large households are unique or rare in the survey counties. To effectively avoid revealing the identity of any large household that may be

the only one or among a very few of its size in its county, data for a number of household members were removed from several households. The number of households' data affected and the number of household members' data removed through this procedure will not be disclosed. The members' data that were removed were not eligible for interview elsewhere in the questionnaire.

Because of these modifications, it may not be possible for analysts to exactly reproduce all of the results in published country reports and indicators.

A. Datasets

In each of the **CSV, SAS, SPSS, and STATA** subfolders, there are 17 datasets for the Northern Kenya 2015 ZOI Interim Survey. In addition to the data files, please note that for the SAS version of the datasets, there is a SAS formats catalog file (containing the value labels for the SAS datasets): **formats.sas7bcat**.

1. **KEN_HOUSEHOLD_PR** – This dataset (n=1,837, vars=102) is a household-level file, and thus contains records for each sampled household with a completed interview in the Northern Kenya 2015 ZOI Interim Survey. The unique identifier for this file is pbs_id, and the dataset includes variables from Modules A, D, and F. Please note that all key household-level derived variables are also included in the household dataset, including, for example, gendered household type (“ghht”) and the household hunger indicator (“hhhungers”). These derived household-level variables can be merged into other datasets as needed using the unique household identifier, pbs_id.
2. **KEN_HHMEMBERS_PR** – This dataset (n=8,555, vars=35) is the household roster for the Northern Kenya 2015 ZOI Interim survey. It has one record per household member from Module C. The unique identifiers for this file are pbs_id + idcode.
3. **KEN_WOMEN_FILE_PR** – This dataset (n=1,476, vars=140) is an individual-level dataset with all women age 15-49 with a completed interview in Module H. The unique identifiers for this file are pbs_id + idcode.
4. **KEN_CHILDREN_FILE_PR** – This dataset (n= 1,435, vars=217) contains records for all children under 5 years of age in the sampled households. It includes data from Module I for children's anthropometry and infant and young child feeding practices. The anthropometry Z-scores were calculated in SAS during the data management process, using the World Health Organization (WHO) SAS “igrowup” package. The unique identifiers for this file are pbs_id + idcode.
5. **KEN_WEAI_PR** – This dataset (n= 1,573, vars=167) is the first of two datasets needed to calculate the WEAI-related measures. It contains Module G data from the primary adult (18+) female decisionmaker within each household (for the sub-sample of households with a primary adult female decisionmaker). The unique identifiers for this file are pbs_id + idcode. (Note that pbs_id alone can uniquely identify records in this WEAI file because, for the interim survey, there was only one person per household [the primary adult female decisionmaker] interviewed for the WEAI module.)
6. **KEN_WEAI_TIMEUSE_PR** – This dataset (n=28,314, vars=113) is the second of the two WEAI datasets used to calculate the WEAI measures. It includes all of the 24-hour time allocation data from Module G6, the time use questionnaire, and thus each respondent in Module G has multiple records, one for each of the 18 time use activities. (28,314 activities ÷ 18 activities = 1,573 Module G respondents with time use data in the file.) The unique identifiers in this long file are pbs_id + idcode + activity.

7. **KEN_WEAI_RECODE_PR** – This dataset (n=1,573, vars=169) contains the analytical variables derived in the WEAI analysis. The unique identifiers in this individual-level file are pbs_id + idcode.
8. **KEN_EXP_PR** – This dataset (n=1,837, vars=66) contains variables from Module E, the Household Consumption Expenditures module used to calculate the poverty and expenditure indicators. It includes household-level derived variables (including the expenditure and poverty indicator variables), as well as variables from sub-Module E6: Housing Expenditures. The unique identifier in this household-level file is pbs_id.
9. **KEN_EXP_E1_PR** – This dataset (n=243,789, vars=28) contains data from sub-Module E1: Food Consumption Over Past 7 Days. Each household with food consumption data has multiple records (for the 133 food items in sub-Module E1). (243,789 records ÷ 133 food items = 1,833 Module E households with sub-Module E1 data.) The unique identifiers in this long file are pbs_id + e101.
10. **KEN_EXP_E2_PR** – This dataset (n=18,300, vars=17) contains data from sub-Module E2: Non-Food Expenditures Over Past 7 Days. Each household with data for non-food expenditures over the past week has multiple records (for the 10 non-food items in sub-Module E2). (18,300 records ÷ 10 non-food items = 1,830 Module E households with sub-Module E2 data.) The unique identifiers in this long file are pbs_id + e201.
11. **KEN_EXP_E3_PR** – This dataset (n=53,070, vars=17) contains data from sub-Module E3: Non-Food Expenditures Over Past One Month. Each household with data for non-food expenditures over the past month has multiple records (for the 29 non-food items in sub-Module E3). (53,070 records ÷ 29 non-food items = 1,830 Module E households with sub-Module E3 data.) The unique identifiers in this long file are pbs_id + e301.
12. **KEN_EXP_E4_PR** – This dataset (n=69,540, vars=17) contains data from sub-Module E4: Non-Food Expenditures Over Past Three Months. Each household with data for non-food expenditures over the past three months has multiple records (for the 38 non-food items in sub-Module E4). (69,540 records ÷ 38 non-food items = 1,830 Module E households with sub-Module E4 data.) The unique identifiers in this long file are pbs_id + e401.
13. **KEN_EXP_E5_PR** – This dataset (n=53,070, vars=18) contains data from sub-Module E5: Non-Food Expenditures Over Past 12 Months. Each household with data for non-food expenditures over the past 12 months has multiple records (for the 29 non-food items in sub-Module E5). (53,070 records ÷ 29 non-food items = 1,830 Module E households with sub-Module E5 data.) The unique identifiers in this long file are pbs_id + e501.
14. **KEN_EXP_E5CONS_PR** – This dataset (n=3,662, vars=15) contain data from sub-Module E5 re. non-food items that may or may not have been purchased (e.g., construction items such as wood poles and thatching grass). Each household with data for these construction items over the past 12 months has multiple records (for the two construction items in sub-Module E5). (3,662 records ÷ 2 construction items = 1,831 Module E households with sub-Module E5 construction data.) The unique identifiers in this long file are pbs_id + e504.
15. **KEN_EXP_E7_PR** – This dataset (n=69,540, vars=23) contains data from sub-Module E7: Durable Goods Expenditures. Each household with data for durable goods expenditures has multiple records (for the 38 durable goods in sub-Module E7). (69,540 records ÷ 38 non-food items = 1,830 Module E households with sub-Module E7 data.) The unique identifiers in this long file are pbs_id + e701.
16. **KEN_Resilience1_PR** – This dataset (n=1,837, vars=84) is a wide file with variables for social capital, adaptive capacity, and asset sales/recovery from the Resilience module in the survey (a portion of Module F). This dataset has one record per household, and contains the derived variables from the Resilience module. The unique identifier for this file is pbs_id.

17. **KEN_Resilience2_PR** – This dataset (n=34,903, vars=13) is a long file with variables for household resources, and their ranking and availability from the Resilience module (a portion of Module F). This dataset includes 19 records (the number of resources asked about in F201) per household (19 x 1,837 households = 34,903 total records), and the unique identifiers are pbs_id + f202_code.

B. Codebook

The Northern Kenya 2015 ZOI Interim Survey codebook includes two worksheets. The “VariableList” worksheet lists the dataset name, variable name, variable label, variable description, variable data type, and (where applicable) value label names. The “Notes” column in this worksheet provides helpful information regarding specific variables. A second worksheet, “Formats_ValueLabels,” provides value label names, values, and value labels for the applicable categorical variables. Both original and derived (analysis) variables are included in this codebook for the Northern Kenya 2015 ZOI Interim Survey.

C. Privacy Assurance Statement (PAS)

A PAS (with informed consent information) for the Northern Kenya 2015 ZOI Interim Survey is also included in this data package. This document presents a general overview of the requirements and process to ensure respondents' privacy and confidentiality prior to public release of survey data. The PAS also includes the informed consent statement from the USAID-approved questionnaire. Informed consent was obtained by trained interviewers from each eligible respondent to the survey.