

User Notes for PMA2018/Kenya Round 7 Household and Female Data, Version 1

Disclaimer: PMA2020 cannot provide in-depth support for data analysis or data related questions, however, to assist the end-user, explanation of some variables is provided below.

General Variables

SIF variables: Date and time variables are provided in both string format and as Stata Internal Format (SIF) values. The variable name of any variable that has been changed into SIF is appended with SIF (e.g. **system_date** and **system_dateSIF**). For all questions requiring a date entry, if the respondent answered either “Do Not Know” or refused to answer the question, the date was recorded as January 1, 2020.

Select multiple variables: Some questions allow for the selection of multiple answers. These variables are in string format and the values are the concatenation of answer choices (e.g. if a household respondent said that they use two sources of water, such as a protected well and rainwater, the value of the observation would read “protected_well rainwater”). Multi-select options are generally, though not always, transformed into binary variables for analysis.

Variable Response Options

Select one: Most select one numeric variables have consistent values for option choices across countries (e.g. **marital_status==1** is equivalent to currently married in all countries). Exceptions include the variables **school**, **fp_provider**, **floor**, **roof**, and **walls**, which have country-specific options and numbering.

Select multiple: Similarly, most select multiple variables have the same response options across countries. Some select multiple variables, however, such as **assets**, have answer options that vary across countries.

See the HHQFQ Master Codebook for complete details on variables and answer choices for each PMA2020 country and survey round.

Specific Variables

EA_ID: The primary sampling unit masked with a random number for anonymity. The same random number is applied to the same EA across multiple rounds.

wealth: Some country datasets include **wealthquintile** while others include **wealthtertile**. The continuous variable **score** is included to allow for construction of various wealth categories.

metainstanceID: metainstanceID is the unique ID generated by ODK for each form submitted to the central server. For PMA2020, the variable metainstanceID is unique for

each household but will be repeated within the household. **memberID** will provide a unique ID for each person within the household.

FQmetainstanceID: FQmetainstanceID is the unique ID generated by ODK for each female form submitted to the central server. For PMA2020, the variable FQmetainstanceID is unique for each female surveyed.

current_recent_methodnum, current_methodnum, recent_methodnum: The numbering scheme for contraceptive methods is consistent across all PMA2020 countries. For example, female sterilization is equal to 1 in every PMA2020 country, whether or not there are any reported uses of female sterilization in the dataset. In some countries, therefore, the numbering will be non-consecutive if some method choices are not selected.

cp, mcp, tcp: Variables that identify current users of any contraceptive method (**cp**), a modern contraceptive method (**mcp**), and a traditional contraceptive method (**tcp**) are included in publicly available datasets so that PMA2020 estimates involving current contraceptive use and method mix can be replicated. Values for these variables are 0 (no) or 1 (yes). PMA2020 codes **cp**, **mcp**, and **tcp** based on the variable **current_methodnum** with the following caveats:

1. Women who report not being a current user of contraception (**current_user=0**), but who report using EC in the past 12 months (**recent_methodnum=8**. emergency) are coded as **cp=1** and **mcp=1**. During analysis, current method is classified as EC in the method mix. The variables **current_methodnum_rc** and **recent_methodnum_rc** reflect this.
2. Women who report using LAM as a current method (**current_methodnum=14**. LAM) must satisfy the three conditions listed below to be coded as **mcp=1**. If any of these conditions are not met, these women are coded as **tcp=1**. During analysis, current method is classified as LAM or traditional method. The variable **current_methodnum_rc** reflects this.
 - a. Less than six months post-partum
 - b. Amenorrheic
 - c. Indicating that they are using LAM with the intention of preventing pregnancy
3. Women who report female sterilization as their first contraceptive method (**first_methodnum=1**. female sterilization), but who do not report currently using female sterilization are coded as **cp=1** and **mcp=1**. During analysis, current method is classified as female sterilization in the method mix. The variable **current_methodnum_rc** reflects this.

GPS Variables

GPS coordinates are not released in this dataset. For some PMA2020 countries a separate dataset of displaced GPS coordinates at the EA level is released.

Notes for Missing Data

In Stata, Missing data is expressed as “.” in the cell. Generally, Stata commands perform computations of any type handle missing data by omitting the row with the missing values.

However, this may vary across commands. PMA does not impute missing values. Missing data in datasets should be studied and/or treated before proceeding to analysis.

Reasons for missing data:

Normal situations:

1. Incomplete forms: If a household, female, or SDP form is not marked as completed (HHQ_result, FRS_result, SDP_result not equal to 1), the observation is likely to miss most of the information. Incomplete forms should not be included in the analysis.
2. Observations that are ineligible for subsequent forms: Only eligible respondents will receive subsequent forms. For example, males and ineligible females will not receive female questionnaires in family planning surveys, hence their observations will have all missing values in female forms.
3. Question not administered due to skip logic: PMA surveys use ODK's skip logic function. The subsequent questions are administered selectively based on the respondent's previous answers. Irrelevant or inapplicable questions are skipped. For example, a woman who is not a contraceptive user will not be asked questions about contraceptive usage subsequently.

Uncommon situations:

1. Lost forms: Due to technical constraints in some challenging data collection areas, forms can be lost in the process of data submission. Although most forms were recoverable, there are occasionally a few that cannot be found. For example, an observation from an eligible woman with completed female form information but missing household form information, or vice versa. These observations may be dropped based on analysis needs.
2. Missing due to incorrect skip logic: PMA surveys were conducted under rigorous quality control. However, in rare cases, there can be incorrect skip logic, which skipped a question that was supposed to be administered, resulting in missing values. These errors are documented in the PMA codebook, which can be downloaded from PMA website. It's not necessary to drop the entire observation since this will likely affect only a few questions.

Distinguish missing data from negative values:

1. -99: No response. The respondent was administered with the question but did not provide an answer. PMA survey requires consent from the respondent and the respondent has the right to refuse to answer any questions at any point. -99 is recorded to reflect that the respondent did not provide an answer to a certain question.
2. -88: Did not know. The respondent consented to answer a specific question but without knowing the answer.
3. -77: Not applicable. The question is administered to the respondent but not applicable to the respondent's situation.

Sampling

PMA2018/Kenya Round 7 used a multi-stage stratified cluster design with urban-rural and selected 11 counties as strata. A sample of 151 enumeration areas (EAs) was drawn by the Kenya National Bureau of Statistics from its master sampling frame. The sample included the two new counties, Kakamega and West Pokot, which were added in 2016, round 5. The same set of enumeration areas used from round 5 were selected, adjacent to the areas enumerated in the first four survey rounds. In each EA, households and private health facilities were listed and mapped, with 42 households randomly selected. Households were surveyed and occupants enumerated. All eligible females age 15 to 49 were contacted and consented for interviews. The final sample included 6,097 households and 5,720 females. Data collection was conducted between November and December 2018

Analytic Sample

PMA2020 analyses include only observations from completed household interviews. The female sample includes only completed female interviews from completed households. The majority of indicators include only de facto women (women who slept in the household the night before). All observations, however, are included in the dataset to allow end users to calculate response rates.

Dataset Version Updates

Any updates made to datasets after their initial release will be documented here.

In January 2017, all previously released datasets were modified as below:

1. The value of **age_at_first_use_children** is 0 for women who have ever used family planning and who have never given birth. Previously, such women had a missing value for age_at_first_use_children.
2. The values for **water_sources_main_drinking** and **water_sources_main_other** equal the value of water_sources_all if a household has one water source. Previously, such households may have had a missing value for these variables.
3. The value for **sanitation_main** equals the value of sanitation_all if a household has one sanitation facility. Previously, such households may have had a missing value for this variable.

All datasets released after January 2017 will have these changes included.

To report errors or inconsistencies:

Please email datamanagement@pma2020.org