

User notes for PMA2018/Nigeria (National) Round 5 Household and Female data, version 2

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.

Generic

SIF variables: Data 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. The values for these variables are the concatenation of answer choices (e.g. if a household respondent said that they use two sources of water, choices a and c, the value of the observation would read “a c”). Multi-select options are generally, though not always, transformed into binary variables for analysis.

Country specific variables: All variables in PMA2020 have consistent values for option choices across countries (e.g. marital_status==1 is equivalent to currently married in all countries) with the exception of the following

1. Livestock questions: The specific livestock options (cow, rabbit etc) vary across countries
2. school: education categories for female schooling vary across countries
3. fp_provider: provider of current or most recent method of family planning vary across countries
4. roof/wall/floor: Household materials vary across country
5. assets: The household assets used to construct wealth scores vary across countries as do the binary variables that are created from the multi-select asset question
6. wealthquintile/wealthtertile: In some countries, wealth quintiles are provided, in other wealth tertiles. The continuous variable score is included to allow for reconstruction of various wealth categories.

Specific Variables

EA_ID/ Cluster_ID: The primary sampling unit. In most countries, EA_ID identifies the primary sampling unit. In Nigeria, enumeration areas were too small to serve as the primary sampling unit. In this case, EAs were clustered and the variable Cluster_ID serves as the primary sampling unit.

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; however, current method is not changed in the data that is publicly available.
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; however, current method is not changed in the data that is publicly available.
 - 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; however, current method is not changed in the data that is publicly available.

GPS Variables

No GPS coordinates for either household or service delivery points will be released for any reason.

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/Nigeria Round 5 used a two-stage cluster design within a sample of seven states. One state was selected using probability proportional to size from among those in each six zones. The seventh state (Kaduna) was allocated to the northwest zone. A total of 302 enumeration areas (EAs) was drawn from the National Population Commission's master sampling frame. Each EA was listed and mapped; 35 to 40 households were randomly selected from each EA. Occupants in selected households were enumerated and eligible females of reproductive age (15-49) were contacted and consented for interview. Data collection was conducted between April and May 2018. The final completed sample included 10,070 households and 11,284 females.

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.

In December 2018, version 2 of this dataset was released. Update from the previous version is as follows:

- Values of the variable **ur** were changed from 0 to 2. Rural.

To report errors or inconsistencies:

Please email datamanagement@pma2020.org